MISSISSIPPI LAW JOURNAL

Volume 80, Number 2 *Winter 2010*

ARTICLES

WOULD AN OPT IN REQUIREMENT FIX THE CLASS ACTION SETTLEMENT? EVIDENCE FROM THE FAIR LABOR STANDARDS ACT — *Charlotte S. Alexander*

FORM AND SUBSTANCE OF THE DOCTRINE OF PIERCING THE CORPORATE VEIL — *Thomas K. Cheng*

AGAINST THE CREATION MYTH OF TEXTUALISM: THEORIES OF CONSTITUTIONAL INTERPRETATION IN THE NINETEENTH CENTURY — *John P. Figura*

THE PEACEMAKER TEST: APPLICATION AND COMPARISON — *Luther T. Munford*

ESSAY

A STATE OF RIGHTLESSNESS: THE EGREGIOUS CASE OF NORTH KOREA — *Morse S. Tan*

COMMENTS

TRANSMUTING MISSISSIPPI'S CURRENT TRANSMUTATION DOCTRINES: ESTABLISHING CLEAR AND CONSISTENT PRECEDENTS TO PROPERTY DIVISION

NURTURING THE UNITED STATES COMMERCIAL SPACE INDUSTRY IN AN INTERNATIONAL WORLD: CONFLICTING STATE, FEDERAL, AND INTERNATIONAL LAW

Digitized by the Internet Archive
in 2012 with funding from
LYRASIS Members and Sloan Foundation

http://archive.org/details/mississippilawjo80char

Become a Friend of the *Mississippi Law Journal*

Today the *Mississippi Law Journal* is funded and operated almost exclusively through student efforts. Increasingly, however, the *Journal* depends on the support of its alumni and friends. You can become a Friend of the *Mississippi Law Journal* by making a tax-deductible contribution to one of the funds listed below. Checks should be made payable to the University of Mississippi Foundation and mailed to Post Office Box 8288, University, Mississippi 38677. *Please indicate on your check to which fund you intend to contribute.*

Robert C. Khayat Scholarship Endowment

The Robert C. Khayat Scholarship Endowment was established in 1995 to honor the dedicated service of Robert C. Khayat, a *Journal* alumnus, professor of law, and former Chancellor of the University. Each year the *Journal* awards the Robert C. Khayat scholarship to a deserving member.

As of June 30, 2010, the corpus of the endowment was $97,117.53. Last year the *Journal* awarded four scholarships in amounts of either $500 or $1000. The *Journal* intends to award larger scholarships as the corpus grows.

Mississippi Law Journal Endowment

The *Mississippi Law Journal* Endowment exists to ensure the *Journal's* future financial stability. As of June 30, 2010, the corpus of the endowment was $270,082.88.

Mississippi Law Journal Operating Fund

All contributions to the *Mississippi Law Journal's* operating fund help the *Journal* to meet its most immediate needs, including office maintenance.

Friends of the *Mississippi Law Journal**

Dr. Guthrie T. Abbott
Mrs. Paula Graves Ardelean
The Honorable Rhesa H. Barksdale
Mr. Ross F. Bass Jr.
Mr. William H. Beck Jr.
Bennett Lotterhos Sulser and Wilson
The Honorable William H. Bizzell
The Honorable Eugene M. Bogen
Mrs. Teri M. Bowie
Mrs. Lorraine Walters Boykin
Mr. John R. Braley IV
Mr. Charles L. Brocato
Mr. Jason W. Burch
Mr. James T. Caldwell
The Honorable Thad Cochran
Mr. Richard G. Cowart
Mr. T. David Cowart
Mr. Frank Crosthwait Jr.
Mr. Arlen B. Coyle
Mr. William M. Deavours
Mrs. Marlane Dove
Mr. W. W. Drinkwater Jr.
Dulin & Dulin Ltd.
Lt. Commander W.J. Dunaway
Dr. Wylene W. Dunbar
Mr. Richard M. Edmonson Sr.
Mr. Herbert C. Ehrhardt
Dr. Martha Jane Ellis
Mr. Roger M. Flynt Jr.
Mr. William M. Gage
Mr. Robert C. Galloway
Mrs. Gretchen L. Gentry
Mr. Eugene B. Gifford Jr.
Mr. Hampton Glover Jr.
Mr. William N. Graham
Mrs. Lillows Ann S. Grayson
Mr. Newt P. Harrison
Mr. Robert H. Hodges
Mr. Jamie G. Houston III
Mr. Gary K. Jones
Mr. W.R. Jones Jr.
Mr. William T. Jones
Mrs. Patricia Fleming Kennedy
Mr. William N. Krucks
The Honorable Tom Stewart Lee
Mrs. Virginia C. Lococo
Mr. Roland Marble Sr.
Mr. M.C. Maxwell
Ms. Lydia H. McIntosh
Mr. Jason N. McNeel
Mr. Joseph R. Meadows
Michael T. & Pauline S. Lewis
Mr. Russell W. Mills
Mr. David W. Mockbee
Mr. H. Dixon Montague
Mr. Dalton M. Mounger
Mrs. Trudy Lumpkin Muench

Mrs. Julie Sneed Muller
Mr. Christopher Tyson Nettles
Mr. James W. O Mara
The Honorable J. D. Orlansky
Mr. Tom Ostenson
Mr. James A. Peden Jr.
Mr. Leroy D. Percy
Mr. Richard T. Phillips
Mr. Rubel L. Phillips
Mr. Benjamin J. Piazza Jr.
Mr. George B. Pickett Jr.
Mr. C. Michael Pumphrey
Mr. Clayton C. Purdom
Mr. Hollaman M. Raney
Mr. Eric Tyson Ray
Mr. Richard Roberts III
Mr. Lucius F. Sams Jr.
Ms. Lillous A. Shoemaker
Mr. George M. Simmerman Jr.
Mr. Jimmy Sledge Jr.
Mr. W. Marion Smith
Mr. Ronald T. Solberg
Mr. Darrell Solomon
Mr. Arthur D. Spratlin Jr.
Mrs. Patricia L. Stevens
Mr. Phineas Stevens
Mr. Samuel James Stigler
Mrs. Molly Fergusson Stuart
Mrs. Nancy H. Stumberg
Ms. Margaret J. Murphy Sweetland
Mr. Frank W. Trapp
Mr. Clint Douglas Vanderver
Mr. J. P. Varner
Mr. Lawrence D. Wade
Mrs. Michelle Clingan Waits
Mr. Tim Walsh
Mr. George D. Warner Jr.
Mr. Irwin G. Waterman
Mr. Louis H. Watson Jr.
Mr. Walker L. Watters
Dr. Robert A. Weems
Mr. Eddie C. Williams
Capt. Ronald J. Wojdyla

** As of November 1, 2010*

The *Mississippi Law Journal* is published four times during the school year. The *Journal*s editorial and business offices may be contacted via U.S. mail at: Post Office Box 849, University, Mississippi 38677.

Manuscripts: Manuscripts should be submitted electronically to **misslj@olemiss.edu**. Manuscripts should be in 12-point type and double-spaced, in Microsoft Word, WordPerfect, or PDF. Citations should conform to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (18th ed. 2005).

Subscriptions: $35.00 per year. Single issue: $12.00. Subscriptions are renewed automatically upon expiration unless the subscriber sends a notice of termination.

Change of address: Send address changes directly to the *Mississippi Law Journal*. Include name, new address (including zip code) and old address. Please notify the *Journal* forty-five days in advance to ensure prompt delivery.

Back Issues: Back issues of the *Mississippi Law Journal* may be ordered directly from William S. Hein & Co., Inc., 1285 Main Street, Buffalo, New York 14029-1987. Orders may also be placed by calling William S. Hein & Co., Inc., at (800) 828-7571, via fax at (716) 883-8100 or email to order@wshein.com. Back issues are available in PDF format through HeinOnline.

The *Mississippi Law Journal* is a member of the National Conference of Law Reviews. To offer its contributors complete freedom, the *Journal* assumes no responsibility for the views expressed herein.

Mississippi Law Journal

Published by Students at
The University of Mississippi School of Law

VOLUME 80 Winter 2010 NUMBER 2

Copyright © 2010, Mississippi Law Journal, Inc.

Editor-in-Chief
PHILIP M. LEVY

Executive Editor
D. ERIC SCHIEFFER

Articles Editors
NATHAN L. BARRETT
MARY MCKENNA
SARAH E. MILLER
JED ROEBUCK
MICHAEL L. TELFORD

Business Manager
JOSEPH SHAYEB

Mississippi Articles Editor
LAUREN OAKS

Mississippi Cases Editor
NEAL WISE

Notes & Comments Editors
JACOB CURTIS
MARK WOODS

Associate Editors

Articles Editors
JOANNA ABE
MARK BOND
BRENNAN P. BREELAND
MARIELLE DIRKX
BRADLEY HIATT
RICHARD J. HUGHES
FRAN MURPHY
ROBERT PARROT
M. MADISON TAYLOR
CORY WILLIAMSON
JAKLYN WRIGLEY
STEPHEN P. HUWE

Mississippi Cases Editors
WILLIAM DRINKWATER
CHRIS HOLLY
BRYAN JONES
MATTHEW ROBERTS
BROOKE TRUSTY

Alumni Chair
LAURA LEE GILLILAND

Symposium Chair
JASON P. VARNADO

Development Editor
PATRICK TIMONY

Managing Editor
THOMAS C. LEITNER, JR

Membership Director
MEGHANN AINSWORTH

Speakers Bureau Chair
MEREDITH BLASINGAME

Faculty Advisors
DEBORAH H. BELL
JOHN CZARNETZKY
ROBERT A. WEEMS

STAFF

Michael Attaway
Alan Baker
Michael Barbee
Brittany Bartley
Parker Berry
Christine Bocek
Jeffrey Brown
Robert Cain
Luke Cantrell
John Carrington
Brent Cole
Will Craven
Josh Daniel
Elliott G. Flaggs
Robert Ford
Wes Francis
Cassi Franks
Andrew S. Harris
April Hendricks
Adria Hertwig
Elizabeth Hyde
Kathleen Ingram
Kate Kinnison Van Namen
Mary Jordan Kirkland
Patrick Lofton
Loren McRae
Bryan Sawyers
Matt Sitton
Morgan Stephens
Austin Stokes
Andrew Tominello

ADMINISTRATIVE OFFICERS OF THE UNIVERSITY OF MISSISSIPPI

DANIEL W. JONES	*Chancellor*
MORRIS H. STOCKS	*Interim Provost and Vice Chancellor for Academic Affairs*
ALICE CLARK	*Vice Chancellor for Research and Sponsored Programs*
LARRY SPARKS	*Vice Chancellor for Administration and Finance*
WENDELL WEAKLEY	*President/CEO, The University of Mississippi Foundation*
ANDREW P. MULLINS, JR.	*Executive Assistant to the Chancellor*
TIMOTHY L. WALSH	*Executive Director of Alumni Affairs*
RICHARD GERSHON	*Dean of the School of Law and Professor of Law*

THE LAW SCHOOL FACULTY

GUTHRIE T. ABBOTT, *Professor Emeritus and Butler, Snow, O'Mara, Stevens and Cannada Lecturer in Law Emeritus*

MICHELE ALEXANDRE, *Assistant Professor of Law*

RICHARD L. BARNES, *Professor of Law and Leonard B. Melvin, Jr. Distinguished Lecturer in Law*

DEBORAH H. BELL, *Professor of Law, Mississippi Defense Lawyers Association Distinguished Lecturer*

WILLIAM W. BERRY, III, *Assistant Professor of Law*

P.J. BLOUNT, *Research Counsel and Instructor*

JOHN R. BRADLEY, *Professor of Law*

PHILLIP W. BROADHEAD, *Director, Clinical Appeals Program and Clinical Professor of Law*

CHARLES BROWER, II, *Croft Associate Professor of International Law and Jessie D. Puckett, Jr. Lecturer in Law*

MERCER E. BULLARD, *Associate Professor of Law*

LARRY BUSH, *Professor Emeritus of Law*

W. TUCKER CARRINGTON, *Director, Mississippi Innocence Project*

DAVID W. CASE, *Associate Professor of Law*

BENJAMIN P. COOPER, *Assistant Professor of Law*

THOMAS K. CLANCY, *Director of the National Center for Justice and the Rule of Law and Research Professor of Law*

GEORGE C. COCHRAN, *Professor of Law*

AARON S. CONDON, *Professor Emeritus of Law*

BENJAMIN COOPER, *Assistant Professor of Law*

JOHN M. CZARNETZKY, *Professor of Law and Mitchell, McNutt, and Sams Lecturer*

DONNA DAVIS, *Associate Professor of Law*

SAMUEL DAVIS, *Jamie L. Witten Chair of Law and Government and Professor of Law*

JASON DERRICK, *Acting Assistant Professor of Legal Writing*

D. MICHAEL FEATHERSTONE, *Professor Emeritus of Law*

MOLLY FERGUSSON, *Acting Assistant Professor of Legal Writing*

Don L. Fruge, *Professor Emeritus and President and CEO Emeritus of the University of Mississippi Foundation*

Joanne I. Gabrynowicz, *Director of the National Remote Sensing, Air and Space Law Center and Research Professor of Law*

Kris L. Gilliland, *Director of Law Library and Associate Professor of Law*

Christopher Green, *Assistant Professor of Law*

Karen O. Green, *Professor of Law and Mississippi Defense Lawyers Association Distinguished Lecturer*

Matthew R. Hall, *Associate Professor of Law and Jessie D. Puckett, Jr., Lecturer*

Desiree Hensley, *Assistant Professor and Director of Civil Legal Clinic*

Michael H. Hoffheimer, *Professor of Law and Mississippi Defense Lawyers Association Distinguished Lecturer*

Michael Johnson, *Counsel for National Programs, National Center for Justice and the Rule of Law and Research Professor*

Robert C. Khayat, *Chancellor Emeritus and Professor Emeritus of Law*

Patricia Krueger, *Acting Assistant Professor of Legal Writing and Director of Academic Excellence Program*

Don Mason, *Associate Director of the National Center for Justice and the Rule of Law and Lecturer*

Gary Myers, *Professor of Law and Ray and Louise Stewart Lecturer*

Jack W. Nowlin, *Associate Professor of Law and Jessie D. Puckett, Jr., Lecturer*

E. Farish Percy, *Associate Professor of Law and Jessie D. Puckett, Jr., Lecturer*

Larry J. Pittman, *Professor of Law and Jessie D. Puckett, Jr., Lecturer*

Lisa S. Roy, *Associate Professor of Law and Jessie D. Puckett, Jr., Lecturer*

Ronald J. Rychlak, *Associate Dean for Academic Affairs, Professor of Law and Mississippi Defense Lawyers Association Lecturer*

Jacqueline Serrao, *Associate Director of the National Remote Sensing, Air and Space Law Center*

Stephanie Showalter, *Director of the National Sea Grant Law Center and Adjunct Professor of Law*

Hans P. Sinha, *Director of Prosecutorial Externship Program and Clinical Professor of Law*

Carolyn Ellis Staton, *Professor Emeritus of Law*

Robert A. Weems, *Professor of Law and Butler, Snow, O'Mara, Stevens and Cannada Lecturer in Law*

C. Jackson Williams, *Director of Legal Writing Program and Acting Assistant Professor of Legal Writing*

MISSISSIPPI LAW JOURNAL
Published by Students at
The University of Mississippi School of Law

VOLUME 80 Winter 2010 NUMBER 2

TABLE OF CONTENTS

ARTICLES

WOULD AN OPT IN REQUIREMENT FIX THE CLASS ACTION SETTLEMENT? EVIDENCE FROM THE FAIR LABOR STANDARDS ACT
.... *Charlotte S. Alexander* 443

FORM AND SUBSTANCE OF THE DOCTRINE OF PIERCING THE CORPORATE VEIL
.... *Thomas K. Cheng* 497

AGAINST THE CREATION MYTH OF TEXTUALISM: THEORIES OF CONSTITUTIONAL INTERPRETATION IN THE NINETEENTH CENTURY
.... *John P. Figura* 587

THE PEACEMAKER TEST: APPLICATION AND COMPARISON
.... *Luther T. Munford* 639

ESSAY

A STATE OF RIGHTLESSNESS: THE EGREGIOUS CASE OF NORTH KOREA
.... *Morse S. Tan* 681

COMMENTS

TRANSMUTING MISSISSIPPI'S CURRENT TRANSMUTATION DOCTRINES: ESTABLISHING CLEAR AND CONSISTENT PRECEDENTS TO PROPERTY DIVISION 709

NURTURING THE UNITED STATES COMMERCIAL SPACE INDUSTRY IN AN INTERNATIONAL WORLD: CONFLICTING STATE, FEDERAL, AND INTERNATIONAL LAW 741

WOULD AN OPT IN REQUIREMENT FIX THE CLASS ACTION SETTLEMENT? EVIDENCE FROM THE FAIR LABOR STANDARDS ACT

Charlotte S. Alexander[*]

I. INTRODUCTION 444
II. RECOVERY-LOSS RATIOS IN RULE 23(B)(3) CLASS ACTION SETTLEMENTS 447
  A. The Absent Plaintiff in Rule 23(b)(3) Class Actions 451
III. THE OPT IN PROPOSAL FOR RULE 23(B)(3) CLASS ACTION REFORM 455
IV. FAIR LABOR STANDARDS ACT CASE STUDY 459
  A. The Design and Structure of the FLSA 459
  B. Case Study Methodology 460
V. RECOVERY-LOSS RATIOS IN SETTLED FLSA COLLECTIVE ACTIONS 462
VI. WHY IS THE OPT IN REQUIREMENT FAILING TO PRODUCE HIGHER RECOVERY-LOSS RATIOS? 465
  A. The Failure of the Opt In Requirement as a Merits Filter 465
  B. The Failure of the Opt In Requirement to Produce Effective Attorney Control 474
VII. CONCLUSION 483
APPENDIX 488

---

[*] Harvard Law School Post-Graduate Research Fellow; Adjunct Professor, Emory University School of Law and Georgia State University College of Law; Deputy Director, National Institute for Teaching Ethics and Professionalism. B.A. 2000, Columbia University; J.D. 2005, Harvard Law School. I thank Harvard Law School and the Skadden Fellowship Foundation for supporting this project. I also thank James Alexander, Abbie Dora Alexander, Aaron Caplan, John Coffee, Clark Cunningham, Gregg Polsky, Michael Jimenez, Gregory Todd Jones, Parina Patel, Al Pearson, Greg Schell, and workshop participants at the Georgia State College of Law for their guidance and suggestions.

I. INTRODUCTION

Class action settlements receive heavy criticism in the popular press and scholarly literature. They are forms of "judicial blackmail";[1] plaintiffs' class action attorneys are "legal vultures."[2] Rule 23(b)(3) class action settlements come under particular fire when attorneys' fees represent the majority of the settlement and the plaintiffs recover only a small fraction of the money they lost.[3] Critics attribute these settlements to two types of unsavory attorney behavior. Either an attorney brings weak claims, which the plaintiffs are forced to settle cheaply, but the attorney still recovers a fee, or the attorney brings strong claims, but "sells out" his or her clients, trading a relatively low plaintiff recovery for relatively high attorneys' fees. The first type of case is often called a "strike suit"; the second is a "sweetheart deal."[4]

Behind these two views of Rule 23(b)(3) class action settlements is an image of the plaintiffs' attorney as a self-interested fee-maximizer.[5] According to critics, Rule 23(b)(3) class action attorneys are particularly likely to engage in fee-maximizing behavior because they are uncontrolled by their clients, who, because of Rule 23(b)(3)'s opt out structure, are effectively absent from the litigation.[6]

---

[1] Martin H. Redish, *Class Actions and the Democratic Difficulty: Rethinking the Intersection of Private Litigation and Public Goals*, 2003 U. CHI. LEGAL F. 71, 71, 73 (2003) ("Much of the scholarly commentary has been highly critical of the modern class action.").

[2] Steve Lohr, *Microsoft Tries to Settle*, N.Y. TIMES, Nov. 25, 2001, at §4, Col. 1.

[3] John C. Coffee, Jr., *Rescuing the Private Attorney General: Why the Model of the Lawyer as Bounty Hunter is Not Working*, 42 MD. L. REV. 215, 225-26 (1983) ("[P]rivate litigated judgments are few, cheap settlements are common, and the typical settlement recovery is below even the level of the compensatory damages alleged by the plaintiffs . . . .").

[4] *See, e.g.,* James D. Cox, *Making Securities Fraud Class Actions Virtuous*, 39 ARIZ. L. REV. 497, 502 (1997) (discussing strike suits); Bruce Hay & David Rosenberg, *"Sweetheart" and "Blackmail" Settlements in Class Actions: Reality and Remedy*, 75 NOTRE DAME L. REV. 1377, 1377-78 (2000) (discussing sweetheart deals). The strike suit characterization, however, has come under attack in the literature as inaccurate. *See, e.g.,* Cox, *supra*, at 502; John C. Coffee, Jr., *Accountability and Competition in Securities Class Actions: Why "Exit" Works Better Than "Voice,"* 30 CARDOZO L. REV. 407, 410 (2008).

[5] Jonathan R. Macey & Geoffrey P. Miller, *The Plaintiffs' Attorney's Role in Class Action and Derivative Litigation: Economic Analysis and Recommendations for Reform*, 58 U. CHI. L. REV. 1, 7 (1991) ("Over the past decade a number of scholars, including most prominently Professor John Coffee, have recognized that the single most salient characteristic of class . . . litigation is the existence of 'entrepreneurial' plaintiffs' attorneys."); Joel Seligman, *Rethinking Private Securities Litigation*, 73 U.

Reformers propose to control attorneys' fee-maximizing behavior by giving these absent plaintiffs a greater role in Rule 23(b)(3) litigation. They would do so by changing the way that plaintiffs join a class action. Presently, all potential plaintiffs who fall within a Rule 23(b)(3) class description are presumed to participate, *en masse*, unless they exclude themselves by opting out. In contrast, in an opt in regime, each individual plaintiff would be required to affirmatively join a case at the outset rather than being included by default.

Reformers believe that this change would control attorneys' fee-maximizing behavior in two ways. First, if a particular case lacks merit, plaintiffs simply will not opt in, depriving attorneys of the chance to bring strike suits designed only to deliver a payment of attorneys' fees.[7] Second, with respect to sweetheart deals, the opt in requirement would produce an autonomous, identifiable group of plaintiffs, rather than a mass of anonymous class members, who would actively participate in settlement negotiations and prevent their attorneys from trading lower plaintiff recoveries for higher attorneys' fees.[8]

CIN. L. REV. 95, 97 (2004) ("In the most extreme form of this critique, the plaintiff's attorney [has been] excoriated as 'entrepreneurial,' or an 'unfaithful champion' of the plaintiff class."). Professor Coffee notes, however, that his examination of attorney behavior "was intended more as a positivist analytic construct than a normative evaluation." John C. Coffee, Jr., *Class Action Accountability: Reconciling Exit, Voice, and Loyalty in Representative Litigation*, 100 COLUM. L. REV. 370, 372 n.3 (2000).

[6] Coffee, *supra* note 3, at 233.

[7] Deborah R. Hensler & Thomas D. Rowe, Jr., *Beyond "It Just Ain't Worth It": Alternative Strategies for Damage Class Action Reform*, 64 LAW & CONTEMP. PROBS. 137, 145 (2001) ("But when the wrongdoing was insignificant or the remedies trivial, few people would come forward. . . . Faced with a higher risk of not covering their costs to pursue a class action, plaintiffs' attorneys would be less likely to file non-meritorious suits . . . in which the likely remedies were trivial."); Edward H. Cooper, *The (Cloudy) Future of Class Actions*, 40 ARIZ. L. REV. 923, 936 (1998) ("If class members opt in at a rate that supports enforcement, well and good. If so few opt in that the litigation founders for want of support, so be it.").

[8] Howard M. Erichson, *Beyond the Class Action: Lawyer Loyalty and Client Autonomy in Non-Class Collective Representation*, 2003 U. CHI. LEGAL F. 519, 523, 554-57

Despite the popularity of this opt in proposal, there has been no empirical study of the workings of an opt in requirement in an actual opt in regime. This is surprising, as at least three major federal statutes have required for decades that plaintiffs opt in, and at least one state's rules of civil procedure permit opt in class actions in certain circumstances.[9] This article begins to fill the gap by examining cases filed and settled under one opt in statute, the Fair Labor Standards Act (FLSA). (This article does not engage with the question of whether Rule 23 reformers are correct in their criticism of attorney behavior but instead uses the critics' own measures to test the assumptions behind their proposed reform.)

The FLSA provides a natural experiment for testing class action reformers' assumptions about the workings of an opt in requirement. If reformers are correct, then one would expect to see plaintiffs in settled FLSA cases recovering a larger percentage of their losses than plaintiffs in settled Rule 23(b)(3) class actions. The FLSA's opt in requirement would filter out weak claims, leaving only those strong cases in which plaintiffs could negotiate favorable settlements. The opt in requirement would also increase client control over their attorneys, producing settlements that maximize plaintiffs' recoveries rather than attorneys' fees.

(2003) (discussing opt in classes as a way to increase plaintiff autonomy; observing that "in non-class litigation . . . litigants and their lawyers make decisions about the conduct of the litigation and the course of settlement negotiations"); John Bronsteen, *Class Action Settlements: An Opt-In Proposal*, 2005 U. ILL. L. REV. 903, 923-25 (2005) (noting that an opt in settlement, like a contract, "empowers . . . the individual herself to decide whether she would benefit from a proposed exchange"); John C. Coffee, Jr., *Litigation Governance: Taking Accountability Seriously* 61 (Columbia Law and Economics Working Paper No. 359, 2010), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id= 1503513 (cited with author's permission) ("[C]lass members who do not opt out in an opt-out class action may still go uncompensated because they often do not file claims post-settlement; in contrast, opt-in class members, having elected to participate, are unlikely to be so indifferent."). Admittedly, this is a rather strange vision of human behavior, assuming that potential plaintiffs will invest significant time, energy, and resources in investigating the lawsuits that they are invited to join and remaining active throughout the course of the litigation.

[9] The federal statutes are the Fair Labor Standards Act, the Equal Pay Act (technically part of the Fair Labor Standards Act), and the Age Discrimination in Employment Act. 29 U.S.C. §§ 201-219 (2006); 29 U.S.C. § 206(d) (2006); 29 U.S.C. §§ 621-634 (2006). The state is Pennsylvania. *See* PA. R.C.P. 1711 (2009).

However, the data presented in this article suggest that the opt in requirement is not functioning in the FLSA context in the way that class action reformers would predict. Plaintiffs not faring much better as a relative matter: the ratio of plaintiffs' recoveries to their losses in FLSA cases is comparable to that in Rule 23(b)(3) class actions. In addition, FLSA opt in rates are low across the board, suggesting that the opt in requirement is not acting as a merits filter. The opt in requirement also does not appear to produce more effective monitoring of attorneys by clients, as the ratio of attorneys' fees to the gross settlement amounts in FLSA cases is greater than the ratio in Rule 23(b)(3) class actions. These observations suggest that the opt in requirement would not be a cure-all for the Rule 23(b)(3) class action. It does not work in FLSA cases in the way that class action reformers predict and likely would not work in any Rule 23(b)(3) case that shares characteristics of FLSA litigation.[10]

This article proceeds as follows: Part II examines data on recovery-loss ratios in the Rule 23(b)(3) context and explores the strike suit and sweetheart deal explanations for class action plaintiffs' relatively low recoveries. Part III explains the opt in proposal for reforming Rule 23(b)(3) class actions and the effects that reformers predict it will have. Part IV describes the design and structure of the FLSA and the methodology for my own study of FLSA settlements. Part V presents my data on recovery-loss ratios in settled FLSA cases. Part VI explores reasons for the apparent failure of the opt in requirement to produce higher FLSA recovery-loss ratios, and Part VII offers final observations on the FLSA and Rule 23(b)(3) class action reform.

II. RECOVERY-LOSS RATIOS IN RULE 23(B)(3) CLASS ACTION SETTLEMENTS

John Coffee, a leading class action scholar, sums up critics' view of class action settlements: "[P]rivate litigated judgments are few, cheap settlements are common, and the typical settlement recovery is below even the level of compensatory damages alleged by the plaintiffs."[11] A number of empirical studies have examined this claim that "cheap settlements are common" in Rule 23(b)(3) class action litigation by comparing plaintiffs' recoveries to their losses—the recovery-loss ratio.[12]

---

[10] *See infra* Part VII for a discussion of shared characteristics.

Figure 1 lists the findings of four studies of recovery-loss ratios. Figure 1 also calculates the median and mean recovery-loss ratio for all four studies taken together and lists the recovery-loss ratio found by the study with the largest sample size.[13] The aggregate median recovery-loss ratio is 5.95%, while the aggregate mean is 13.57%. The study with the largest sample size reported a recovery-loss ratio of either 15.41% or 25.63%, depending on how the plaintiffs' losses were calculated. Because the studies had different sample sizes, employed different methodologies, and examined different types of cases in different years, it would be inappropriate merely to group them together and form conclusions based on the aggregate median or mean or only to look at the largest study. Instead, considering the aggregate median, the mean, and the findings of the study with the largest sample size together may present a more complete picture of the data. Whichever statistic is used, the plaintiffs in these studies rarely recovered much more than twenty-five cents on the dollar, or 25%, of the money they claimed to have lost due to the defendants' misconduct.

---

[11] Coffee, *supra* note 3, at 225-26.

[12] Professor Coffee also alludes to the plaintiffs' compensatory damages as a possible alternative benchmark for assessing whether the plaintiffs' recovery is adequate. *Id.* at 226. However, this particular comparison does not appear to have been studied empirically.

[13] The median is the midpoint of the data; the mean is the average.

Figure 1: Class Action Studies: Recovery-Loss Ratios

Authors	**Years Studied**	**Case Type**	**Sample Size**	**Recovery-Loss Ratio**
Bajaj et al.[14]	1991-1999	Securities	1,300 (approx.)	0.12% - 30.69% or 0.22% - 51.04%
Buckberg et al.[15]	2002-2005	Securities	957	2.5% - 2.9%
Cox et al.[16]	1993-2005	Securities	736	1% - 9%
Hensler et al.[17]	1989-1996	Consumer and mass tort	10	13% - 61%

Largest sample size (Bajaj et al.)	1991-1999	Securities	1,300 (approx.)	15.41% - 25.63% [median]
Aggregate mean	n/a	n/a	n/a	13.57%
Aggregate median	n/a	n/a	n/a	5.95%

Of course, many variables influence the size of a plaintiff's recovery, and a recovery-loss ratio of 25% may not, in and of itself, be an indication that the plaintiff's attorney is maximizing his or her fee at the plaintiff's expense through a strike suit or sweetheart deal. In addition, the very nature of the recovery-loss ratio calculation may be misleading, as the "loss" figure is generated by the plaintiffs, who have every incentive to be optimistic in their loss estimates. Nevertheless, the 25% figure is useful for the purposes of this article as a benchmark for assessing the workings of an opt in requirement: class action reformers view a 25% recovery-loss ratio as unacceptably low and believe that this figure would increase in an opt in regime.[18]

---

[14] Mukesh Bajaj, Sumon C. Mazumdar & Atulya Sarin, *Empirical Analysis: Securities Class Action Settlements*, 43 SANTA CLARA L. REV. 1001, 1021 tbl. 14 (2002) (finding recovery-loss ratios between 0.12% and 30.69% using one method of calculating the plaintiffs' losses and between 0.22% and 51.04% using another method of loss calculation).

[15] Elaine Buckberg, et al., Nat'l Econ. Research Assocs., *Recent Trends in Shareholder Class Action Litigation: Are WorldCom and Enron the New Standard?* 2, 6 (July 2005), *cited with approval in* Coffee, *supra* note 8, at 29 n.69.

[16] James D. Cox, Randall S. Thomas & Lynn Bai, *There Are Plaintiffs and . . . There Are Plaintiffs: An Empirical Analysis of Securities Class Action Settlements*, 61 VAND. L. REV. 355, 367, 375 tbl. 6 (2008) (showing ratio of settlement amount to provable losses; reporting median ratios of 0.01 to 0.09).

[17] DEBORAH R. HENSLER ET AL., RAND INSTITUTE FOR CIVIL JUSTICE, CLASS ACTION DILEMMAS: PURSUING PUBLIC GOALS FOR PRIVATE GAIN 140-41, 425-27 (2000).

Critics contend that low recovery-loss ratios are the result of either strike suits or sweetheart deals. The "strike suit" explanation holds that suits are "brought for their nuisance value rather than for any harm actually suffered by the members of the class [and therefore that] class members recover extremely small amounts."[19] Strike suits are characterized by "small cash settlements that represent a small percentage of [the plaintiffs'] damages," i.e., recoveries that are low both as an absolute matter and in comparison to the losses the plaintiffs claimed.[20]

In the "sweetheart deal" view, "a substantial portion of the settlement [is] paid to the plaintiff's attorney" rather than to the plaintiffs, and the attorneys' fees can be as large as or larger than the plaintiffs' recovery.[21] In an example provided by Professor Coffee, a self-interested plaintiffs' attorney in this situation "would prefer a $500,000 settlement out of which a $300,000 award of attorneys' fees would be paid, to a $1,000,000 recovery out of which only a $200,000 fee would be paid," even though the plaintiffs would recover four times as much under the second scenario.[22]

---

[18] Curiously, studies of the recovery-loss ratio in Rule 23 class actions rarely, if ever, compare that ratio to any baseline in non-class cases. This raises the question of how class action critics determine whether any particular recovery-loss ratio is "low" or "high."

[19] Cox, *supra* note 4, at 502.

[20] Cox et al., *supra* note 16, at 383.

[21] Seligman, *supra* note 5, at 97.

[22] Coffee, *supra* note 3, at 243; *id.* at 232-33 ("To say this is not to claim that plaintiffs' attorneys systematically subordinate the class recovery to their own fee, but it is to say that the plaintiff's attorney is subject to a serious conflict of interest—one that can distort the settlement process and reduce the deterrent effect of private litigation—whenever the determination of the fee award is not made a sufficiently direct function of the size of the recovery so as to align the interests of the private enforcer with those of the class he purports to represent."); Susan P. Koniak & George M. Cohen, *In Hell There Will Be Lawyers Without Clients or Law*, 30 HOFSTRA L. REV. 129, 145 (2001) ("Lawyers have a keen interest in the size of their fees, while '[c]lients are interested primarily in the size of their recovery.' To the degree lawyers dominate the lawyer-client relationship, there is a danger that they will 'engage in conduct that increases their fees . . . at the expense of the client's recovery.'" (alteration in original) (quoting Susan P. Koniak & George M. Cohen, *Under Cloak of Settlement*, 82 VA. L. REV. 1051, 1103-04 (1996)).

These two explanations have in common an image of the plaintiffs' attorney as functionally independent from his or her clients. Critics charge that a lack of affirmative involvement by Rule 23(b)(3) class members from the litigation's beginning to end liberates the class action attorney from the plaintiff class, allowing the attorney to engage in fee-maximizing behavior. This absent plaintiff class is the target of the opt in reform.

A. The Absent Plaintiff in Rule 23(b)(3) Class Actions

Class actions lawsuits under Rule 23(b)(3) begin with a named plaintiff and an attorney. In some cases, the named plaintiff approaches the attorney and seeks representation. In others, the attorney identifies a possible claim and recruits a named plaintiff who holds that claim.[23] Regardless of how a case is initiated, if it is not dismissed on a defense motion, the litigation proceeds to class certification, in which the plaintiffs propose a class definition and seek to represent all class members. If the court grants certification, notice is sent to all class members informing them of the litigation and their right to opt out. If the case settles, class members may receive a second notice of the opportunity to object before the court accepts the proposed settlement. If class members do not opt out or object, they become bound by the terms of the settlement and eligible to collect their share of the plaintiffs' recovery.[24]

Critics charge that this procedure results in "[t]he effective absence of the client" in Rule 23(b)(3) litigation.[25] Studies of Rule 23(b)(3) class actions have shown that vanishingly few plaintiffs do opt out. Figure 2 summarizes opt out rates observed in four class action studies. As Figure 2 shows, the aggregate median opt out rate for the four studies combined was only 0.15% and the aggregate mean was 0.3%, while the study with the largest sample size reported an opt out rate of 0.1%. The corresponding plaintiff participation rates were 99.85%, 99.7%, and 99.9%, respectively.

---

[23] Macey & Miller, *supra* note 5, at 99 ("That such solicitation occurs on a regular basis is patently obvious.").

[24] *See* FED. R. CIV. P. 23(c)(2)(B); FED. R. CIV. P. 23(e)(1). Those class members who do exclude themselves have the option to file their own lawsuits, separate from the class action, but may not recover as part of any judgment awarded to or settlement collected by the class.

[25] Coffee, *supra* note 3, at 233.

Figure 2: Class Action Studies: Plaintiff Opt Out and Participation Rates

Authors	**Years Studied**	**Case Type**	**Sample Size**	**Plaintiff Opt Out Rate**	**Plaintiff Participation Rate**
Bertelsen et al.[26]	1966-1972	Various	36	0%-10%	100%-90%
Eisenberg & Miller[27]	1993-2003	Various	143	0.1%	99.9%
Hensler et al.[28]	1989-1996	Consumer and mass tort	10	0.65%	99.35%
Willging et al.[29]	1992-1994	Various	84	0.1%-0.2%	99.9%-99.8%

Largest sample size (Eisenberg & Miller)	1993-2003	Various	143	0.1%	99.9%
Aggregate mean	n/a	n/a	n/a	0.3%	99.7%
Aggregate median	n/a	n/a	n/a	0.15%	99.85%

---

[26] Bruce I. Bertelsen, Mary S. Calfee & Gerald W. Connor, Note, *The Rule 23(b)(3) Class Action: An Empirical Study*, 62 GEO. L. J. 1123, 1126-27, 1161 (1974) (finding that in 31 of 36 cases studied, 10% or less of the class chose to opt out, and in 15 of those cases, no class member opted out).

[27] Theodore Eisenberg & Geoffrey Miller, *The Role of Opt-Outs and Objectors in Class Action Litigation: Theoretical and Empirical Issues*, 57 VAND. L. REV. 1529, 1532, 1546 (2004) ("The median percentage of class members opting out, in the 143 cases in which the opinion reveals both the number of opt-outs and the number of class members, is a mere 0.1 percent.").

[28] HENSLER ET AL., *supra* note 17, at 325-26 (reporting in one case study 25 opt-outs out of approximately 3800 class members, or 0.65%).

[29] Thomas E. Willging, Laural L. Hooper & Robert J. Niemic, *Empirical Study of Class Actions in Four Federal District Courts: Final Report to the Advisory Committee on Civil Rules* 4, 52, 134 fig. 43 (Federal Judicial Center 1996) (finding in 84 cases that "the median percentage of members who opted out was either 0.1% or 0.2% of the total membership of the class and 75% of the opt-out cases had 1.2% or fewer class members opt out.").

Regardless of which figure is used, these studies reveal that opt out rates in Rule 23(b)(3) class actions are exceedingly low and that plaintiff participation rates are exceedingly high. However, though large numbers of plaintiffs appear to be "participating" in Rule 23(b)(3) class actions, it is likely that class members' nearly universal failure to opt out is a sign of passivity rather than an affirmative decision to stay in the litigation. As Theodore Eisenberg and Geoffrey Miller comment, "[c]ommon sense indicates that apathy, not decision, is the basis for inaction."[30] This is an example of what Cass Sunstein has called a "sticky" default option.[31] As Sunstein and Richard Thaler summarize, supported by social science research in a variety of arenas:

> [M]any people will take whatever option requires the least effort, or the path of least resistance. . . . [F]or a given choice there is a default option—an option that will obtain if the chooser does nothing—then we can expect a large number of people to end up with that option, whether or not it is good for them.[32]

In Rule 23(b)(3) cases, the default is set at participation and the path of least resistance for class members is not to opt out. As a result, though class actions appear to have extremely high plaintiff participation rates, the vast majority of those plaintiffs have never made any affirmative effort to take part in the litigation.

In fact, a Rule 23(b)(3) class action is structured such that plaintiffs have little incentive to act and every incentive to remain passive. According to prominent class action observers Jonathan Macey and Geoffrey Miller:

> Upon receiving and reading notice of the suit, the typical plaintiff in a large-scale, small-stakes class action has a choice between two courses of action. She can do nothing, in

---

[30] Eisenberg & Miller, *supra* note 27, at 1561.

[31] *See generally* Cass R. Sunstein, *Switching the Default Rule*, 77 N.Y.U. L. REV. 106 (2002) (discussing "sticky" default rules).

[32] RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE 83 (Yale Univ. Press 2008).

> which case she will receive a check in the mail if the suit is successful and will incur no costs if the suit fails. Or she can go to the trouble of opting out of the action, in which case she will receive nothing whether or not the suit is successful. Such a decision is not hard to make. Nearly everyone who understands the nature of this choice will elect to do nothing and thereby remain part of the class action.[33]

The rational choice for the typical plaintiff, with low damages and little to lose, is thus to remain in the case, passively waiting for a check in the mail.

This lack of active involvement carries through to the end of the litigation as well. Though Rule 23(b)(3) class action plaintiffs have the right to object to any proposed settlement, studies have found that the incidence of any form of objecting behavior is "infrequent" and "trivial."[34] Moreover, even when class members could put money directly into their pockets by taking action, studies show that they do not. A 2000 study by the RAND Institute for Civil Justice of Rule 23(b)(3) consumer and mass tort class actions found that while plaintiffs were promised between 13% and 61% of their losses in settlement, when class members were required to file a claim to receive their settlement monies, "the fraction of compensation funds actually disbursed to class members was modest to negligible."[35]

---

[33] Macey & Miller, *supra* note 5, at 28 (citations omitted); *see also* Cooper, *supra* note 7, at 936 ("Inertia, the complexity of class notices, and the widespread fear of any entanglement with legal proceedings will lead many reluctant class members to forgo the opportunity to opt out . . . ."). Those few plaintiffs who do choose to opt out might be those with the highest individual claims and the greatest incentive to engage with the case. Their withdrawal from the class action leaves behind only those plaintiffs with small claims and little reason or inclination to participate actively in the litigation or to monitor their attorneys. John C. Coffee, Jr., *Rethinking the Class Action: A Policy Primer on Reform*, 62 IND. L. J. 625, 633 (1987).

[34] Willging et al., *supra* note 29, at 57; Eisenberg & Miller, *supra* note 27, at 1546 ("The median percentage of class members objecting is zero in the 205 cases with the necessary data. Opt-out and objector percentages are thus trivial.").

[35] HENSLER ET AL., *supra* note 17, at 459; *see also* Samuel Issacharoff & Geoffrey P. Miller, *Will Aggregate Litigation Come to Europe?*, 62 VAND. L. REV. 179, 205 (2009) (observing that claiming rates are quite low and class action plaintiffs often leave significant sums "on the table"); James D. Cox & Randall S. Thomas, *Letting Billions Slip Through Your Fingers: Empirical Evidence and Legal Implications of the Failure of*

Critics blame the opt out structure of Rule 23(b)(3) for this sort of plaintiff passivity. In this view, it is "the effective absence of the client" that enables attorneys' fee-maximizing strike suits and sweetheart deals and ultimately produces low recovery-loss ratios. According to Professor Coffee, "the parties can find a variety of means by which to trade a low settlement for a high attorney's fee, once the client becomes only a distant bystander to the litigation."[36]

III. THE OPT IN PROPOSAL FOR RULE 23(B)(3) CLASS ACTION REFORM

A prominent reform proposal in response to this rather dismal portrayal of Rule 23(b)(3) class action settlements has been to change the way that plaintiffs enter a class action from

---

Financial Institutions to Participate in Securities Class Action Settlements, 58 STAN. L. REV. 411, 412 (2005) (also observing the significant sums left "on the settlement table").

[36] Coffee, *supra* note 3, at 232; Koniak & Cohen, *supra* note 22, at 146 ("[C]lient monitoring of lawyer performance is effectively unavailable in almost all class actions."). Not only do absent class members fail to monitor their attorneys, but arguably, neither does the lead plaintiff, who, though nominally the representative of the absent class, is at times recruited by the attorney in the first place and is therefore essentially "captured." Further, though Rules 23(e) and 23(h) require court approval of all class action settlements, including an assessment of the settlement's fairness to absent class members, commentators have long complained that many courts "rubber stamp" settlements to which the parties have agreed. John C. Coffee, Jr., *Class Wars: The Dilemma of the Mass Tort Class Action*, 95 COLUM. L. REV. 1343, 1348 (1995); Koniak & Cohen, *supra* note 22, at 149-54 (discussing pressures on judges to approve settlement agreements); Susan P. Koniak, *Feasting While the Widow Weeps:* Georgine v. Amchem Products, Inc., 80 CORNELL L. REV. 1045, 1120 (1995) ("[O]ne must search long and hard to find any court opinion that has disapproved a class settlement on the ground that it was collusive. Thus, the 'absence of collusion' standard [for court approval of class action settlement agreements] is meaningless as now applied.") (citation omitted). Research has confirmed these complaints: a study in the Second Circuit has shown, for example, that even district courts that were specifically admonished by the appellate court to monitor attorneys' fee awards closely and keep fees in check did not achieve much of a reduction of fee awards. Theodore Eisenberg, Geoffrey Miller & Michael A. Perino, *A New Look at Judicial Impact: Attorneys' Fees in Securities Class Actions After* Goldberger v. Integrated Resources, Inc., 29 WASH. U. J.L. & POL'Y 5, 16, 18-19 (2009) (studying 717 securities class action fee awards in district courts in the Second Circuit Court of Appeals before and after appeals court decision admonishing district court judges to scrutinize fee awards more closely and finding that fee awards decreased from the 28.35%–30% range only to the 26.03%–27.25% range).

opt out to opt in, to give the absent clients a greater role in the litigation. A brief history of the Rule 23(b)(3) opt out mechanism is useful in understanding this proposed reform.

When the Federal Rules of Civil Procedure were first promulgated in 1938, Rule 23 permitted three categories of class action: "true," "hybrid," and "spurious."[37] The spurious class action had an opt in requirement, and only those plaintiffs who intervened in the case would be bound by and able to recover under the judgment.[38] The landscape changed in 1966 when the opt in spurious class action was replaced by today's opt out Rule 23(b)(3).

Class action reformers today want to change the landscape again, to switch the default back to some semblance of the intervention or opt in spurious model. This proposal has surfaced regularly in the literature and in public debate over class action reform. As early as 1972, the first major study of class actions after the 1966 creation of Rule 23(b)(3) proposed two opt in models: that courts "either (1) require individuals to affirmatively 'opt in' by a specified date and thus become class members, or (2) allow individuals to become class members through their failure to request exclusion but require such class members to file a proof of claim by a specified date in order to benefit from any ultimate recovery."[39] The Advisory Committee on Civil Rules again considered the opt in reform during its deliberations in 1996,[40] and the opt in proposal has appeared in various forms in law review articles and public discourse until the present day.[41]

---

[37] The "true" class action comprised claims with common questions of law or fact. This form of action allowed parties whose claims would otherwise be subject to mandatory joinder, but whose numbers made joinder impracticable, to proceed collectively, and the result was binding on absent parties. The "hybrid" class action comprised claims that, while not all concerning the same questions of law or fact, all related to specific property. Judgment in hybrid class actions was also binding on absent class members.

[38] *See generally* Benjamin Kaplan, *Continuing Work of the Civil Committee: 1966 Amendments of the Federal Rules of Civil Procedure (I)*, 81 HARV. L. REV. 356 (1967); Stephen C. Yeazell, *The Past and Future of Defendant and Settlement Classes in Collective Litigation*, 39 ARIZ. L. REV. 687, 696 n.41 (1997). This sort of action was called "spurious" because it was not in any true sense a class action, but rather a mechanism for the permissive joinder of individual plaintiffs. *Id.* at 385.

[39] AMERICAN COLLEGE OF TRIAL LAWYERS, REPORT AND RECOMMENDATIONS OF THE SPECIAL COMMITTEE ON RULE 23 OF THE FEDERAL RULES OF CIVIL PROCEDURE 32 (1972).

Reformers believe that the opt in proposal would address both strike suits and sweetheart deals—and ultimately increase plaintiffs' recovery-loss ratios—by controlling attorneys' fee-maximizing behavior. First, the number of plaintiffs who opt in would act as a check on attorneys' ability to bring non-meritorious strike suits. Potential plaintiffs would choose to join only those strong cases that would result in high recoveries. Attorneys would learn to file only meritorious cases, because non-meritorious cases would draw such low plaintiff participation that they would not be worth the investment of attorney resources and time.[42] Second, the opt in requirement would produce a plaintiff class that would be more autonomous and engaged than Rule 23(b)(3) plaintiffs. Each plaintiff would be named and identifiable, able to weigh in on settlement offers and participate meaningfully in litigation decisions. These active opt in plaintiffs would closely monitor their attorneys and prevent them from trading lower plaintiff recoveries for higher attorneys' fees, the sort of fee-maximizing behavior that characterizes sweetheart deals.[43]

---

[40] Memorandum from Patrick E. Higginbotham, Chair, Advisory Comm. on Civil Rules, to Honorable Alicemarie H. Stotler, Chair, Standing Comm. on Rules of Practice and Procedure 35 (May 17, 1996), *available at* http://207.41.14.183/rules/Reports/CV5-1996.pdf ("This draft would have collapsed the categorical distinctions now observed between subdivision (b) (1), (b) (2), and (b) (3) classes; authorized the court to permit or deny opting out of any class action; created an opt-in class provision; specifically governed notice requirements for (b)(1) and (b)(2) classes; and made many other changes, many of them independently significant.").

[41] *See, e.g.,* AMERICAN COLLEGE OF TRIAL LAWYERS, *supra* note 39, at 32-34; Cooper, *supra* note 7, at 935-37; HENSLER ET AL., *supra* note 17, at 476; *Public Hearing: Proposed Amendments to the Federal Rules of Civil Procedure Rule 23* (N.D. Cal. Jan. 17, 1997) (testimony of Alfred W. Cortese, Jr.), *reprinted in* 3 WORKING PAPERS OF THE ADVISORY COMMITTEE ON CIVIL RULES ON PROPOSED AMENDMENTS TO CIVIL RULE 23 (1997) at 280-86; 4 WORKING PAPERS OF THE ADVISORY COMMITTEE ON CIVIL RULES ON PROPOSED AMENDMENTS TO CIVIL RULE 23 (1997) 388-92 (written statement of Alfred W. Cortese, Jr.); Hensler & Rowe, *supra* note 7, at 145 n.26 (quoting Paul Niemeyer, Speech to the National Press Club (May 20, 1998)); Redish, *supra* note 1, at 130-32; Bronsteen, *supra* note 8, passim; John Bronsteen & Owen Fiss, *The Class Action Rule*, 78 NOTRE DAME L. REV. 1419, 1446 (2003). Professors Coffee, Issacharoff, and Miller also discuss the advantages of the opt in structure. *See* Coffee, *supra* note 15, at 56-68; Issacharoff & Miller, *supra* note 35, at 202-08.

[42] Hensler & Rowe, *supra* note 7, at 145; Cooper, *supra* note 7, at 936.

If these predictions are accurate, then one would expect to see higher recovery-loss ratios in settled FLSA cases—an opt in regime—than the 25% ratio observed in Rule 23(b)(3) class action settlements. As set out in detail below, however, the median recovery-loss ratio in my FLSA study was only 28%. This is so, I suggest, because the opt in requirement does not function as a merits filter in the way that class action reformers predict: plaintiffs' decisions about whether to opt in are influenced by a variety of complex factors and are not merely a comment on a case's merits. Cases filed under the FLSA are therefore probably a mix of strong and weak, just like their Rule 23 counterparts.

In addition, the opt in requirement does not appear to control attorneys' fees as expected by class action reformers. The median ratio of attorneys' fees to gross settlement amounts in the FLSA context was 50%, compared to a 30%–32% ratio in the Rule 23(b)(3) context. In fact, the FLSA's opt in requirement may actually *decrease* plaintiffs' recoveries in relation to attorneys' fees: instead of engaging in the expensive notice and opt in process, many plaintiffs' attorneys have adopted a high-volume, small-case, quick-settlement approach that compensates them for their investment in the case but delivers only a minimal settlement to the plaintiffs. The remainder of this article explains these findings, showing ultimately that the opt in requirement does not behave in the FLSA context in the ways hoped for or expected by class action reformers.

---

[43] Another set of proposals has advocated limiting class action plaintiffs' ability to opt out. Though approaching the problem from a different perspective, this proposal is in fact quite similar to the opt in proposal: the common goal is to bring plaintiffs into the litigation and keep them there, to allow them to police their attorneys and participate more meaningfully in the lawsuit. *See, e.g.*, Coffee, *supra* note 4, at 419; Richard A. Nagareda, *The Preexistence Principle and the Structure of the Class Action*, 103 COLUM. L. REV. 149, 227 (2003) (discussing mandatory classing).

IV. FAIR LABOR STANDARDS ACT CASE STUDY

A. The Design and Structure of the FLSA

The FLSA's opt in requirement dates from 1947. Previously, the FLSA had contained its own collective action mechanism—separate from the class action structure under the Federal Rules—that permitted an agent to bring claims on behalf of the entire plaintiff class. In 1947, Congress revised the FLSA to require each individual plaintiff affirmatively to opt in by "giv[ing] his consent in writing to become such a party and [filing] such consent . . . with the court."[44]

When the Federal Rules of Civil Procedure were amended in 1966 to create the present-day Rule 23(b)(3) opt out class action, the Advisory Committee made clear that the changes did not affect opt in FLSA collective actions. Since 1966, therefore, the opt in FLSA collective action has been carved out from the Federal Rules' opt out class action structure.

Though the FLSA is silent on the mechanics of group representation, courts have constructed a two-step "certification" process for FLSA collective actions. After the original plaintiffs file the complaint, if the court grants conditional collective action certification, the plaintiffs issue a court-approved opt in notice to all potential plaintiffs using defendant-provided names and addresses. The court sets a deadline by which new plaintiffs must file their consent forms in order to join the case. After the opt in period ends and discovery closes, the defendant may move for decertification, arguing that the plaintiffs are not, in fact, similarly situated or that insufficient plaintiffs have opted in to justify collective action treatment.[45] If the case is not decertified, it proceeds to settlement or judgment on behalf of only those plaintiffs who have opted in.

---

[44] 29 U.S.C. § 216(b) (2006) (requiring individual plaintiffs to opt in); *see also id.* at §§ 251-262 (amending the FLSA through the Portal-to-Portal Act). In fact, even before the FLSA was amended in 1947, many courts had already imposed an opt in requirement on FLSA collective actions by treating them as spurious Rule 23 class actions. Marc Linder, *Class Struggle at the Door: The Origins of the Portal-to-Portal Act of 1947*, 39 BUFF. L. REV. 53, 168-69 (1991) (discussing courts' "limit[ing] participation in FLSA actions to named plaintiffs, intervenors, and consenters who joined the action before the trial on the merits"). When Congress created the FLSA's opt in requirement in 1947, it was essentially codifying what courts had already been doing in practice.

The FLSA's opt in requirement is thus essentially an enactment of the opt in reform proposed by Rule 23(b)(3) class action critics. Collective action lawsuits under the FLSA can therefore serve as a natural experiment on the effects of the opt in requirement on strike suits and sweetheart deals and, ultimately, on the percentage of the plaintiffs' losses that they are able to recover.

B. Case Study Methodology

My study of the FLSA's opt in requirement uses as its data set FLSA cases filed in the U.S. District Court for the Southern District of Florida between 2003 and 2008. I chose the Southern District of Florida because, as Figure 3 shows, it has been the primary site for FLSA case filings over the past five years accounting for, in various years, between a quarter and a third of all FLSA cases filed in federal court nationwide.[46] The Southern District is thus the site where FLSA litigation is happening in this country, providing an example of an opt in regime in full flower. (Though I do not know the extent to which the Southern District's FLSA docket is unique compared to the FLSA dockets of other federal district courts, because the FLSA is a federal statute, one can expect some consistency in FLSA litigation characteristics in courts across the country.)

---

[45] *See generally* Hoffman-La Roche, Inc., v. Sperling, 493 U.S. 165 (1989) (describing collective action certification procedure). Though *Hoffman-La Roche* was brought under the Age Discrimination in Employment Act (ADEA), ADEA and FLSA cases use the same court-created collective action certification procedure and the same opt in requirement at 29 U.S.C. § 216(b).

[46] ADMINISTRATIVE OFFICE OF THE U.S. COURTS, FEDERAL JUDICIAL CASELOAD STATISTICS: 2003-2008, tbl. C-2, *available at* http://www.uscourts.gov/Statistics/FederalJudicialCaseloadStatistics.aspx (last visited Oct. 19, 2010); *see also* Denise Martin et al., Nat'l Econ. Research Assocs., *Class Certification in Wage and Hour Litigation: What Can We Learn from Statistics?* 3 (Nov. 9, 2009) (finding that 45% of FLSA cases in 2007 and 2008 filed in U.S. district courts were filed in Florida).

Figure 3: FLSA Cases Filed in the Southern District of Florida and Nationwide (2003-2008)

Year	**FLSA Cases Filed, S.D. Fla.**	**FLSA Cases Filed, Nationwide**	**Percentage of FLSA Cases Filed in the S.D. Fla.**
2003	903	4,055	22
2004	1,040	3,426	30
2005	1,234	3,464	36
2006	1,197	4,389	27
2007	1,378	6,786	20
2008	1,154	5,302	22
Total:	6,906	27,422	25

Using FLSA data from the Southern District, I examined three variables: recovery-loss ratios in settled FLSA collective actions, the rate at which plaintiffs opted into FLSA collective actions, and the ratio of attorneys' fees to gross settlement amounts in settled FLSA collective actions.[47] I employed the following methodology. Using the online Public Access to Court Electronic Records (PACER) database, I identified 7,259 FLSA cases that were filed in all divisions of the Southern District in the years 2003 through 2008. I then removed 208 duplicate records that were created when a case was closed and later reopened within the study period. Finally, I removed 145 cases that were still in litigation as of August 15, 2009, producing a set of 6,906 unique closed cases. From this set, I took a random sample of 364 cases, or just over 5%. I reviewed each of these 364 cases individually and identified 250 that were filed as collective actions, rather than individual lawsuits. These 250 closed collective actions provide the basis for my study.

---

[47] Because of limitations on my data, I was unable to examine all three data points at once—the recovery-loss ratio, fee-settlement ratio, and opt in rate—in any one case or group of cases.

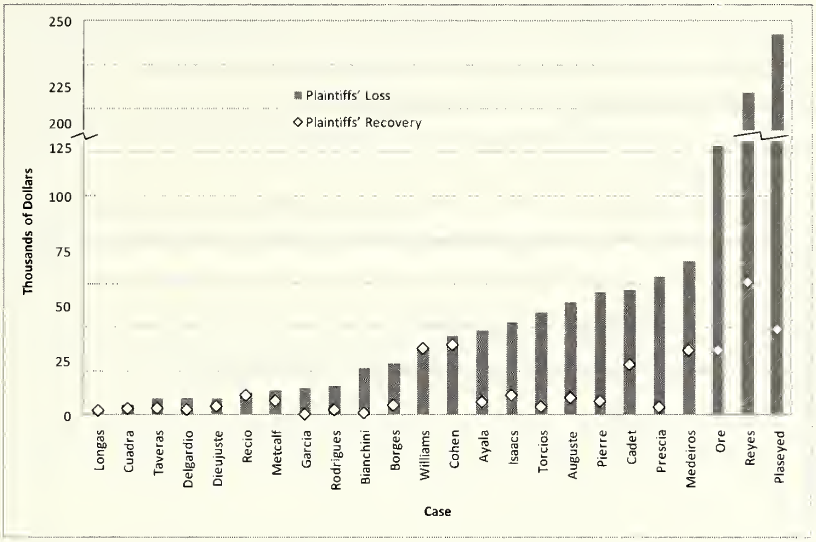
V. RECOVERY-LOSS RATIOS IN SETTLED FLSA COLLECTIVE ACTIONS

To examine recovery-loss ratios in FLSA settlements—the key variable studied in the class action context—I reviewed each of the 250 FLSA collective actions in my sample and identified 218 that ended in settlement before trial. Of these settled collective actions, data on both the plaintiffs' initial estimates of their losses and their ultimate recovery were available in twenty-four cases.[48] Figure 4 shows the plaintiffs' recovery in relation to their loss in each case. The dollar figures, as well as recovery-loss ratios, for each case are available in Figure 9 in the Appendix. Taken together, the plaintiffs in these cases recovered between 4% and 100% of their claimed losses, with a median of 28%. (In reporting the results of my study, I use the median to describe the middle or central tendency of the data rather than the mean, which is influenced by particularly high or low "outlier" data points.) In other words, the typical plaintiff recovered twenty-eight cents for each dollar of loss that he or she suffered.

---

[48] Plaintiffs' estimates of their losses are found in the complaint, in the settlement agreement, or in the "statement of claim" required by some Southern District of Florida judges to be filed at the outset of FLSA cases. In the cases for which complete data was not available, the plaintiffs did not estimate their losses in any document, or the parties filed a sealed settlement agreement (or did not file one at all), or both. I acknowledge that the 25 cases for which I have data might differ from the 193 cases for which I do not have data in substantive ways that could undermine my conclusions. Moreover, as with Rule 23(b)(3) class action recovery-loss ratios, the value of this data is somewhat limited by the necessity of using plaintiffs' own estimates of the size of their losses, which are likely skewed high in the plaintiffs' favor.

Figure 4: Plaintiffs' Losses and Recoveries in Settled FLSA Collective Actions (Sorted by Plaintiffs' Loss)



In the Rule 23(b)(3) context, critics argue that a recovery-loss ratio of approximately 25%, or twenty-five cents of recovery for each dollar of loss, is an indication of strike suit behavior or sweetheart deals: that the plaintiffs' claims lacked merit or that the plaintiffs' attorneys settled for less than their clients' claimed losses in order to extract attorneys' fees. While the FLSA median recovery-loss ratio of 28% exceeds the 25% benchmark figure from the Rule 23(b)(3) context, the margin is quite small, raising questions about the ability of the opt in requirement to increase plaintiffs' recoveries substantially.

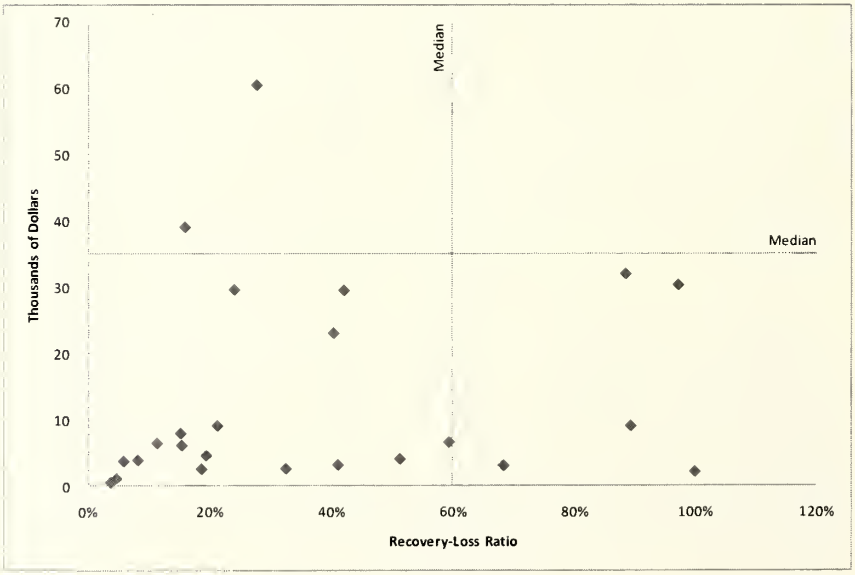
Class action critics also look to the plaintiffs' compensatory damages in assessing the adequacy of the plaintiffs' recovery, observing that Rule 23(b)(3) strike suits and sweetheart deals result in recoveries that are "below even the level of compensatory damages alleged by the plaintiffs."[49] Many of the plaintiffs in my FLSA study also appear to have collected less than their compensatory damages. In FLSA cases, the plaintiffs' total damages consist of their compensatory damages plus an equal amount in liquidated damages. If a FLSA plaintiff recovers less than half of his or her total damages, it follows that he or she has not recovered any liquidated damages and only a portion of his or her compensatory damages. The FLSA plaintiffs in my study recovered 28% of their losses—less than half of their total damages and less than their compensatory damages alone.

---

[49] Coffee, *supra* note 3, at 226.

Furthermore, Figure 5 presents data that suggest that, despite the opt in requirement, some FLSA settlements may exhibit strike suit characteristics: "small cash settlements that represent a small percentage of [the plaintiffs'] damages."[50] Figure 5 plots plaintiffs' recoveries against their recovery-loss ratios in each of the twenty-four cases. Most cases cluster in the lower left quadrant, meaning that the size of the recovery was small both as an absolute matter—the y axis—and in comparison to the size of the plaintiffs' loss—the x axis.

Figure 5: Plaintiffs' Recoveries Versus Recovery-Loss Ratios in Settled FLSA Collective Actions



---

[50] Cox et al., *supra* note 16, at 383.

Thus, the opt in requirement does not appear to have the substantial reformative effect on recovery-loss ratios that class action critics would predict. Plaintiffs' recoveries in the FLSA settlements in my data set were generally small, both absolutely and relatively, and the 28% median FLSA recovery-loss ratio is comparable to the 25% ratio found in the Rule 23(b)(3) class action context. The question, then, is why the opt in requirement is failing to produce the higher recovery-loss ratios in settled FLSA collective actions that class action reformers would expect.[51]

VI. WHY IS THE OPT IN REQUIREMENT FAILING TO PRODUCE HIGHER RECOVERY-LOSS RATIOS?

The opt in requirement could be failing to produce higher recovery-loss ratios in settled FLSA collective actions for at least two reasons. First, the opt in requirement does not seem to be performing the merits filtering function—the proposed solution to the strike suit problem—that class action reformers envision. Second, the plaintiffs who join FLSA cases via the opt in process do not seem to monitor or control their attorneys—the proposed solution to the sweetheart deal problem—sufficiently to prevent the trading of recoveries for fees.

A. *The Failure of the Opt In Requirement as a Merits Filter*

In determining whether an opt in requirement is effectively filtering out non-meritorious cases, one must first identify an appropriate benchmark. What sort of opt in rate would class action reformers accept as a proxy for a meritorious case? Here, it seems safe to assume that reformers would prefer an opt in rate as close to 100% as possible. If the opt in rate is a stand-in for the strength of a case and a signal to plaintiffs' lawyers that the case is worth pursuing, then an indication that this screening mechanism is functioning as predicted would be near-100% opt in rates in the cases that are actually filed.[52]

---

[51] Because I did not control for FLSA-specific variables in my study, I acknowledge the possibility that characteristics of FLSA litigation or of FLSA plaintiffs, instead of or in addition to the failure of the opt in requirement, could generate the results I observe.

To calculate the rate at which plaintiffs opted into the FLSA cases in my study, I screened each of the 250 collective actions in my data set to find those in which collective action certification was granted, notice was issued, and potential plaintiffs had the full court-ordered period to opt in before the case was settled or otherwise concluded.[53] Identifying these particular cases, which I call "complete" collective actions, was necessary to avoid examining cases that were terminated during the opt in period, thereby cutting off plaintiffs' opportunity to join and skewing the opt in rate. Of the 250 cases in my data set, 1.2%, or three, were complete collective actions. In order to identify additional cases, I went back to the full set of 6,906 unique closed cases. Multiplying the same 1.2% figure from my sample by 6,906, I predicted that approximately fifty-seven cases would be complete collective actions. I then reviewed randomized groups of cases, 200 at a time, from the 6,906-case set until I had identified fifty-seven complete collective actions.

Next, I compared the number of notices that were sent to potential plaintiffs in each of the fifty-seven cases with the number of plaintiffs who actually opted in. Each case docket reported the number of plaintiffs who opted in; I gathered data on the number of notices sent, i.e., the number of potential plaintiffs, by contacting the plaintiffs' lawyers in each of the fifty-seven cases. I was able to gather data from the plaintiffs' attorneys in thirty-eight of the fifty-seven cases.

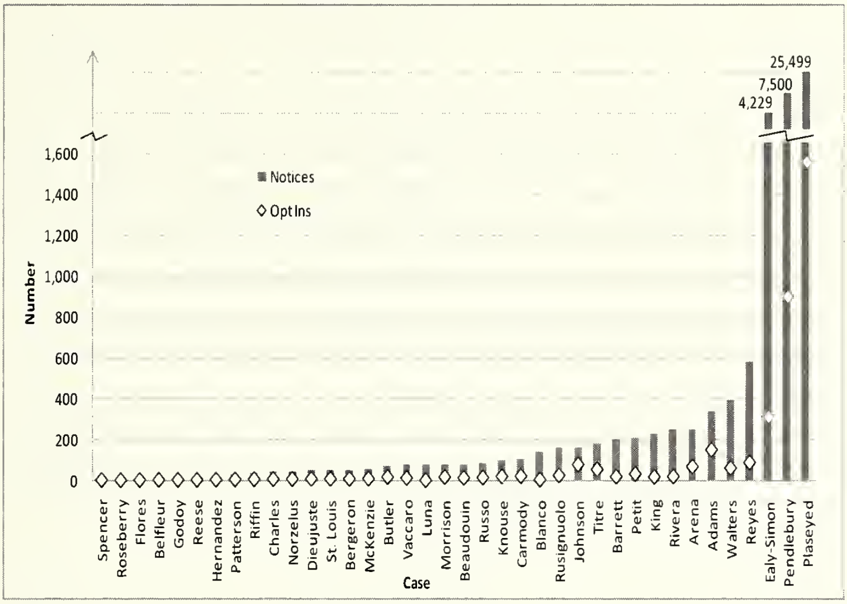
Figure 6 shows the number of notices sent in each of the thirty-eight cases and the number of plaintiffs who opted in. Plaintiff opt in rates in these cases ranged from 0% to 48%, with a median of 15%. The number of notices and opt ins, the opt in rates, the length of the opt in period, and the form of notice are listed in Figure 10 in the Appendix.

---

[52] Switching the focus to the Rule 23(b)(3) context, for example, one can assume that class action reformers would not be bothered by the 99.9% of class members who do not opt out if those plaintiffs could be assumed to be making an affirmative judgment about the merits of the class action and choosing to remain in the litigation rather than "participating" merely by remaining passive.

[53] It was not necessary to limit this group of cases only to those that settled, as in Figures 4 and 9, because the opt in rate can be calculated regardless of the case outcome, as long as the case lasts through the end of the opt in period.

Figure 6: Notices and Opt Ins in Complete FLSA Collective Actions (Sorted by Number of Notices)



The 15% opt in rate observed by my study squares with anecdotal descriptions of opt in rates offered by plaintiffs' attorneys in the Southern District of Florida. One highly experienced FLSA attorney reported opt in rates of between 5% and 90% in the thirty or so cases he had handled in the past decade, with the average opt in rate falling between 13% and 25%.[54] Another veteran FLSA attorney who practices in the Southern District commented that the typical opt in rate is between 10% and 20%, but often even lower.[55] Industry publications have also reported lawyers' estimates of opt in rates ranging from

---

[54] Interview by author (Oct. 9, 2009).

[55] Interviews by author (Dec. 23, 2009).

15% to 30%.[56] Finally, the one analysis of FLSA opt in rates that I have located in the academic literature echoes my Southern District of Florida findings: a 2008 study of twenty-one FLSA collective actions filed in federal courts around the country calculated the average opt in rate to be 15.71%.[57]

If the opt in requirement were functioning as class action reformers would expect, then the median opt in rate for these settled FLSA cases would be substantially closer to the benchmark of 100%. Every case filed would be meritorious, as the only cases that attorneys would accept would be the strong ones that could attract a high opt in rate. The much lower 15% opt in rate found by my study could mean one of two things. On the one hand, if the opt in rate is in fact a proxy for the merits (if plaintiffs flock to strong cases and avoid weak ones), then all cases in my study are strike suits, filed by attorneys who have ignored the "message" of a low 15% opt in rate. On the other hand, the opt in requirement might not actually function as a merits filter, and something else might explain the low 15% opt in rate. Because it simply does not seem plausible that *every* FLSA case in my randomly chosen data set is a strike suit, I suggest that the latter explanation is more accurate.[58]

---

[56] Victoria Roberts, *Attorneys Discuss Strategies for Bringing, Defending FLSA Collective Action Lawsuits*, DAILY LAB. REP. (BNA), Aug. 13, 2002, at C-1 (reporting estimated opt in rates of 15%–30% in FLSA collective actions); Victoria Roberts, *Attorneys Explore Reasons for Surge in Wage and Hour Lawsuits, Offer Strategies*, DAILY LAB. REP. (BNA), Dec. 12, 2002, at C-3 (reporting estimated opt in rate of 20%).

[57] Andrew C. Brunsden, *Hybrid Class Actions, Dual Certification, and Wage Law Enforcement in Federal Courts*, 29 BERKELEY J. EMP. & LAB. L. 269, 292-94 (2008). Though Professors Issacharoff and Miller state that "[e]vidence from reported cases suggests that participation rates under [the FLSA, Age Discrimination in Employment Act, and Equal Pay Act] are around fifty percent," they cite the opt in rates in only two reported FLSA cases to support this proposition. Issacharoff & Miller, *supra* note 35, at 204 & n.71.

[58] Opinions on whether or not most FLSA cases are meritorious tend to be dictated by the alignment of the speaker with the plaintiffs' or defense bar. Plaintiffs' lawyers contend that recent FLSA litigation has merit: because of the economic downturn, employers are squeezed, and they, in turn, squeeze their employees, committing FLSA violations that result in legitimate lawsuits to vindicate those workers' rights. Victoria Roberts, *Attorneys Explore Reasons for Surge in Wage and Hour Lawsuits, Offer Strategies*, DAILY LAB. REP. (BNA), Dec. 12, 2002, at C-2. Members of the defense bar, in contrast, see FLSA litigation as generally meritless, driven by plaintiffs' attorneys intent on lining their pockets. These criticisms are remarkably similar to criticisms of

The FLSA's opt in requirement does not seem to act as a merits filter in the way that class action reformers expect because of flaws in reformers' assumptions about plaintiff behavior. Class action reformers assume that the only factor in a potential plaintiff's opt in decision is the strength of the merits. Statements such as, "[l]arge numbers of class members would come forward . . . when there was a widely shared perception that the alleged wrongdoing was significant and that the potential remedies were worth pursuing" express this assumption.[59] Though the merits might be one factor that potential FLSA plaintiffs consider in deciding whether to opt in, the picture is much more complex in reality, with a variety of forces keeping potential plaintiffs passive.

As an initial matter, the very same apathy and inertia described by Professors Sunstein and Thaler that prevent Rule 23(b)(3) class members from opting out probably also keep potential FLSA plaintiffs from opting in. In the Rule 23(b)(3) context, a class member who is apathetic or inertia-bound is nevertheless included in the litigation. In FLSA cases, the default is the opposite, and those potential plaintiffs who do not take action are excluded. Apathy and inertia may be a particular problem in cases in which the per-plaintiff recoveries are quite small: potential plaintiffs may decide that the time and effort required to complete and return an opt in form are not worth their potential (low) future recovery. These potential plaintiffs have not asserted an opinion about the merits by failing to opt in; they have merely chosen the "option [that] requires the least effort, or the path of least resistance."[60]

Rule 23(b)(3) class action attorneys as self-interested fee-maximizers. *The Wage and Hour Minefield: Some Words of Wisdom for Employers.*, 15 METROPOLITAN CORPORATE COUNSEL 41 (Nov. 2007) ("Colombo: I am seeing a flood of lawsuits in the Southern District of Florida, and at least half of them are overtime-related cases. I attribute this to the attorney fee provision in the FLSA, which is certainly an incentive for lawyers to bring these types of claims, even if the amounts at issue are small."); Simon J. Nadel, *As Overtime Suits Renew FLSA Debate, Attorneys Advise Learning the Wage Laws*, 3 CLASS ACTION LITIG. REP. (BNA) 514 (Aug. 9, 2002) (quoting representative from public policy advocacy organization as saying: "For attorneys who land such a [FLSA] case . . . 'that's a pretty good pay day.'").

[59] Hensler & Rowe, *supra* note 7, at 145.

[60] THALER & SUNSTEIN, *supra* note 32, at 83.

Before a potential plaintiff even reaches this decision point, however, he or she must first receive, read, and understand an opt in notice. My data suggest that some plaintiffs' attorneys in FLSA cases are foregoing notice entirely, meaning that many of the plaintiffs who did not opt in *may not have even known about the case*. All 250 FLSA cases in my data set were initially filed as collective actions, but the plaintiffs moved for collective action certification and notice in only twenty-eight cases, or 11%. The remaining 222 cases were litigated on behalf of only the named plaintiffs, or on behalf of the named plaintiffs and those few additional plaintiffs who had opted in, presumably having heard of the litigation through informal channels. A low opt in rate in these circumstances can hardly be read as an indication of weak merits.[61]

Even if notice does issue, if it is made by mail, which in the cases listed in Figure 10 was by far the most common form of notice delivery, potential FLSA plaintiffs may not receive it. Many FLSA plaintiffs, suing for their unpaid minimum or overtime wages, are low wage earners. Low wage earners, as a class, move frequently, and so may not receive a notice mailed to their last known address as maintained by their employer.[62] This problem is magnified when the plaintiffs are transnational migrants, such as farmworkers and other seasonal employees who return to their countries of origin during the offseason. Attorneys in the United States who represent migrant farmworkers, for example, report that more than one third of the notices they send to potential plaintiffs' employer-provided addresses in Mexico are returned as undeliverable.[63] (Though some commentators have suggested that the advent of the internet and email and the possibility of an electronic notice process could ease the difficulties associated with notice delivery, the familiar problem of the "digital divide" might mean that electronically-delivered notices would not reach low-wage FLSA plaintiffs).[64]

---

[61] An additional barrier to notice might be courts' denials of plaintiffs' requests for collective action certification and notice. There were insufficient motions for collective action certification in my data set to study this question. However, plaintiffs' attorneys in the Southern District of Florida commented in interviews that some judges delay ruling on collective action motions until nearly the end of the case, when the parties are already close to settlement, functionally thwarting the notice process. Interview by author (Dec. 21, 2009); *see also* Daniel C. Lopez, *Collective Confusion: FLSA Collective Actions, Rule 23 Class Actions, and the Rules Enabling Act*, 61 HASTINGS L.J. 275, 308 (2009) (observing that FLSA collective actions "are only fully certified in two narrow instances: single-location production plants or a group of employees with nearly identical job classifications.").

[62] Linder, *supra* note 44, at 175 (noting that the "peripatetic living conditions" of many FLSA plaintiffs "frequently disable them from opting in to [collective] actions"); Craig Becker & Paul Strauss, *Representing Low-Wage Workers in the Absence of a Class: The Peculiar Case of Section 16 of the Fair Labor Standards Act and the Underenforcement of Minimum Labor Standards*, 92 MINN. L. REV. 1317, 1326 (2008) (noting that "the rate of relocation is highest among people with the lowest income").

Potential FLSA plaintiffs also may be unable to read and understand the complicated legal language of an opt in notice, as low-wage workers complete fewer years of schooling on average and are less likely than other workers to speak and read English fluently.[65] Moreover, many people who receive an unsolicited notice from a court about a legal proceeding may be, at best, confused, and, at worst, frightened. "[S]ensible people do not sign legal documents they have not requested and send them to a court or lawyers they do not know and with whom they have never spoken."[66] The problem of a lack of understanding may be exacerbated when courts set very short time periods in which potential opt in plaintiffs must decipher or translate the notice, consult an attorney or other advisor, call the plaintiffs' counsel, and decide whether to join. Courts allowed potential opt in plaintiffs thirty or fewer days to make this complicated decision in one-third of the cases listed in Figure 10. Even for those plaintiffs who receive an opt in notice, therefore, language and comprehension barriers, rather than an assessment of the merits, might drive the opt in decision.

---

[63] Interview by author (Jan. 4, 2010).

[64] E-mail from John C. Coffee, Adolf A. Berle Professor of Law, Columbia Law School, to Charlotte Alexander, Harvard Law School Post-Graduate Research Fellow (Jan. 26, 2010) (on file with author).

[65] Bureau of Labor Statistics, *Characteristics of Minimum Wage Workers*, tbl. 6 (2008), *available at* http://www.bls.gov/cps/minwage2008.htm (last visited Oct. 20, 2010) (reporting that 578,000 of 2,226,000, or 26%, of hourly workers over 16 years old who made the minimum wage or below did not have a high school diploma); Randy Capps et al., *A Profile of the Low-Wage Immigrant Workforce*, URBAN INSTITUTE, Nov. 2003, at 1, *available at* http://www.urban.org/publications/310880.html (last visited Oct. 20, 2010) ("Immigrants are 11 percent of all U.S. residents, but 14 percent of all workers and 20 percent of low-wage workers. . . . Nearly two-thirds of low-wage immigrant workers do not speak English proficiently, and most of these workers have had little formal education.").

[66] Becker & Strauss, *supra* note 62, at 1328.

Yet even if potential plaintiffs do receive, understand, and overcome their apathy or inertia vis-à-vis a FLSA opt in notice, they *still* might make a rational choice not to opt into even the most meritorious case, as many would be required to sue their employers and risk retaliation.[67] Many low wage FLSA plaintiffs can be easily replaced by other workers, a fact that unscrupulous employers emphasize through threats. A 2004 study of the U.S. meat and poultry industries, for example, quoted a worker describing such threats:

> Tyson always gets rid of workers who protest or who speak up for others. When they jumped from thirty-two chickens a minute to forty-two, a lot of people protested. The company came right out and asked who the leaders were. Then they fired them. They told us "If you don't like it, there's the door. *There's another eight hundred applicants waiting to take your job.*"[68]

FLSA caselaw is replete with retaliation lawsuits,[69] and there are doubtless many more retaliation stories that are never the subject of litigation but nonetheless exert a powerful silencing effect.[70] Potential plaintiffs' acute knowledge of their own expendability may leave them with no choice but *not* to opt into a FLSA lawsuit. The situation may be even worse for those potential FLSA plaintiffs who are undocumented or who hold temporary employment visas issued in the name of their employer.[71] For these workers, the potential cost of suing is not only the loss of a job and the associated income, but loss of their legal status in and removal from the United States.[72]

---

[67] My data do not reveal how many FLSA plaintiffs were current employees of the defendants they sued and how many were former employees. Former employees may not have the same fear of retaliation.

[68] HUMAN RIGHTS WATCH, BLOOD, SWEAT, AND FEAR: WORKERS' RIGHTS IN U.S. MEAT AND POULTRY PLANTS 80 (2004) (emphasis added).

[69] *See, e.g.*, Laura Wides-Munoz, *Guest Farmworkers Say Visa System Means Abuse*, THE NEWS-PRESS (Fort Myers, FL), June 3, 2007, at Metro, B7 (describing lawsuit in which twenty workers initially joined, but sixteen withdrew due to fear of retaliation); *see also* Recinos-Recinos v. Express Forestry, Inc., No. 05-1355, 2006 U.S. Dist. LEXIS 2510 (E.D. La. Jan. 23, 2006) (granting protective order prohibiting defendants from contacting potential opt in plaintiffs in FLSA collective action upon evidence that defendants' agents had visited workers' families at their homes in El Salvador and threatened them). Though the FLSA prohibits retaliation, the statute's protections are triggered only *after* a worker has suffered a demotion, transfer, or termination in retaliation for making a FLSA claim. The availability of a possible future claim for damages is therefore cold comfort to many low-wage workers who are facing the decision whether to opt into a FLSA collective action.

Thus, the opt in rate in any given FLSA case likely reflects a complex array of factors, including potential plaintiffs' apathy or inertia, lack of knowledge or understanding, and fear of retaliation.[73] This is a much more nuanced picture of plaintiff behavior than the simple merits-based decision that class action reformers imagine plaintiffs would make in an opt in regime. While Rule 23(b)(3) class action plaintiffs may not share all of the characteristics of potential FLSA plaintiffs, to the extent that they are similar (beset by apathy or inertia; limited in language or comprehension skills; required to sue their employers; vulnerable to retaliation; or unable to expend the time and resources required to opt in), their opt in decisions would be similarly complex, not merely a comment on the merits of the class actions they are invited to join. Given this complexity, one cannot assume that the opt in requirement would function as a merits proxy or a merits filter, screening out strike suits and allowing only the meritorious cases to proceed.

---

[70] David Weil & Amanda Pyles, *Why Complain? Complaints, Compliance, and the Problem of Enforcement in the U.S. Workplace*, 27 COMP. LAB. L. & POL'Y J. 59, 83 (2005) (noting that "being fired is widely perceived to be a consequence of exercising certain workplace rights."). The same Human Rights Watch report, for example, describes a memo issued by a manager at a poultry processing plant in Arkansas in response to rumors that workers were going to protest their wages and working conditions. The memo promised explicitly, "If myself or any of the other management team members hear you say this or another employee tells us about this and it can be backed up, you will no longer work here." HUMAN RIGHTS WATCH, *supra* note 68, at 80.

[71] A 2010 report by the U.S. Department of Homeland Security estimated the number of "unauthorized immigrants," ages eighteen or over, in the U.S. population to be 10.8 million, large numbers of whom are employed as low-wage workers. MICHAEL HOEFER, NANCY RYTINA & BRYAN C. BAKER, U.S. DEP'T OF HOMELAND SECURITY, OFFICE OF IMMIGRATION STATISTICS, ESTIMATES OF THE UNAUTHORIZED IMMIGRANT POPULATION RESIDING IN THE UNITED STATES: JANUARY 2009 1 (2010); U.S. BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, FOREIGN-BORN WORKERS: LABOR FORCE CHARACTERISTICS IN 2008, USDL 09-0303 (2009) (summarizing demographic data). Foreign workers may also be employed in low-wage jobs in the United States with temporary "guestworker" visas under the H-2A and H-2B programs. 8 U.S.C. §§ 1101(a)(15)(H)(ii)(a)-(b) (2006).

[72] Relatively more secure and powerful FLSA plaintiffs may be insulated from the threat of retaliation faced by low wage minimum wage and overtime plaintiffs. For example, in a collection of FLSA overtime cases on behalf of financial advisors, securities brokers, and loan officers provided by Professor Coffee, opt in rates ranged from 48% to 84%. E-mail from John C. Coffee, Adolf A. Burle Professor of Law, Colombia Law School, to Charlotte Alexander, Harvard Law School Post-Graduate Research Fellow (May 7, 2010) (on file with author). These relatively powerful FLSA plaintiffs were probably less affected by the array of factors that keep the more typical FLSA plaintiff from opting in. As suggested by Professor Coffee, the way in which plaintiffs' attorneys are compensated may also play a role. If attorneys are paid as a percentage of the plaintiffs' actual recovery, rather than by the hour or as a percentage of a "common fund" against which plaintiffs must make claims, it is in their interest to seek out plaintiffs (within ethical bounds) and encourage as high an opt in rate as possible. *Id.*

B. The Failure of the Opt In Requirement to Produce Effective Attorney Control

Just as the FLSA's opt in requirement does not filter cases based on the merits, it also does not seem to produce closer attorney monitoring by active clients. Whether the plaintiffs are attempting to monitor their attorneys and failing, or whether they are failing to monitor at all, the opt in requirement does not seem to prevent the trading of recoveries for fees and may in fact *decrease* plaintiffs' recoveries, thereby increasing the portion of the settlement "pie" that is allotted to attorneys' fees.

In the Rule 23(b)(3) class action context, critics often point to the ratio of attorneys' fees to the gross settlement amount in identifying sweetheart deals. In the scenario offered by Professor Coffee, for example, the fee-maximizing attorney preferred "a $500,000 settlement out of which a $300,000 award of attorneys' fees would be paid, to a $1,000,000 recovery out of which only a $200,000 fee would be paid."[74] The fee-settlement ratio in the first instance, which benefits the attorney, is 60%; in the second instance, which benefits the plaintiffs, it is 20%.

---

[73] These observations are not evidence of "paternalism," as one commentator has suggested, but rather the realities of low-wage work and workers in the United States labor market. Redish, *supra* note 1, at 101.

[74] Coffee, *supra* note 3, at 243; *id.* at 232 ("To say this is not to claim that plaintiffs' attorneys systematically subordinate the class recovery to their own fee, but it is to say that the plaintiff's attorney is subject to a serious conflict of interest—one that can

As Figure 7 shows, studies of class actions have found actual fee-settlement ratios ranging from 7% to a whopping 1050%. These studies' aggregate median fee-settlement ratio is 30%, and the aggregate mean is 32%. The study with the largest sample size reports a fee-settlement ratio of 30%. Put another way, of every dollar the defendants paid in settling these class actions, they paid between thirty and thirty-two cents to the attorneys in fees and the remaining seventy to sixty-eight cents to the plaintiffs.

Figure 7: Class Action Studies: Fee-Settlement Ratios

Authors	**Years Studied**	**Case Type**	**Sample Size**	**Fee-Settlement ratio**
Bertelsen et al.[75]	1966-1972	Various	13	33.3% - 50+%
Coffee[76]	Various	Various	Multiple studies aggregated	30%
Cole[77]	1939	Securities	54	20%

---

distort the settlement process and reduce the deterrent effect of private litigation—whenever the determination of the fee award is not made a sufficiently direct function of the size of the recovery so as to align the interests of the private enforcer with those of the class he purports to represent."); Koniak & Cohen, *supra* note 22, at 145 ("Lawyers have a keen interest in the size of their fees, while '[c]lients are interested primarily in the size of their recovery.' To the degree lawyers dominate the lawyer-client relationship, there is a danger that they will 'engage in conduct that increases their fees . . . at the expense of the client's recovery.'") (alteration in original) (quoting Susan P. Koniak & George M. Cohen, *Under Cloak of Settlement*, 82 VA. L. REV. 1051, 1103-04 (1996)).

[75] Bertelsen et al., *supra* note 26, at 1164-65 (finding that the fee-settlement ratio did not exceed one-third in nine of eleven cases, was less than 25% in five of the eleven cases, and exceeded 50% in only two of the eleven cases).

[76] Coffee, *supra* note 15, at 25 n.59 (collecting studies showing fee-settlement percentages between 10% and 30% and up to 32%).

[77] Douglas G. Cole, *Counsel Fees in Stockholders' Derivative and Class Actions–Hornstein Revisited*, 6 U. RICH. L. REV. 259, 273 (1972) (describing a 1939 study of 54 securities class actions that found fee-settlement ratios "ranging between 20[%] and 33.3[%] with the resultant average being slightly over 20[%]") (citation omitted).

Authors	**Years Studied**	**Case Type**	**Sample Size**	**Fee-Settlement ratio**
Eisenberg & Miller[78]	1993-2003	Various	159	25%
Garth et al.[79]	1979-1984	Various	14	7% - 1050%
Jones[80]	1971-1978	Securities	13	16.2%
Kennedy[81]	1966-1973	Securities	13	19%
Willging et al.[82]	1992-1994	Various	68	33.3%

Largest sample size (Coffee)	Various	Various	Multiple studies aggregated	30%
Aggregate mean	n/a	n/a	n/a	32%
Aggregate median	n/a	n/a	n/a	30%

A 30%–32% fee-settlement ratio does not appear overly objectionable and is even slightly lower than the one-third contingency fee ratio that judges commonly use to assess fee petitions and weed out those that compensate the plaintiffs' attorneys excessively.[83] It is likely the fee-settlement ratios at the high end of the spectrum—1050%, for example—that drive critics' characterizations of Rule 23(b)(3) class action settlements as "sweetheart deals."[84] Nevertheless, if class action reformers are correct that the opt in requirement would produce active client monitoring of what they portray as fee-maximizing behavior by attorneys, then one would expect fee-settlement ratios in an opt in regime to be lower than the 30%–32% figure in the Rule 23(b)(3) context.

---

[78] Eisenberg & Miller, *supra* note 27, at 1546 tbl. 1, 1553 (finding that the median "fee as a percent of recovery" for 159 class actions was 25%).

[79] Bryant Garth, Ilene H. Nagel & S. Jay Plagler, *The Institution of the Private Attorney General: Perspectives from an Empirical Study of Class Action Litigation*, 61 S. CAL. L. REV. 353, 355, 371-74 (1988) (finding fee-settlement ratios in fourteen cases ranging from 7% to 1050%, producing a median of 57%).

[80] Thomas M. Jones, *An Empirical Examination of the Resolution of Shareholder Derivative and Class Action Lawsuits*, 60 B.U. L. REV. 542, 564-67 (1980) (reporting fee-settlement ratios in 13 cases up to 40.2% with an average (mean) of 16.2%).

[81] John E. Kennedy, *Securities Class and Derivative Actions in the United States District Court for the Northern District of Texas: An Empirical Study*, 14 HOUS. L. REV. 769, 819 (1977) (finding that attorneys' fees in thirteen settlements "generally constituted from 5% to 33% of the recovery," producing a median of 19%).

[82] Willging et al., *supra* note 29, at 69 (noting in 68 class actions that the fee-settlement ratio "infrequently exceeded the 33.3% contingency fee rate"); *id.* at 147 fig. 68.

[83] HENSLER ET AL., *supra* note 17, at 435-36.

[84] Rule 23(b)(3) class actions that are handled by nonprofit organizations may present a different profile with respect to attorneys' fees. *See, e.g.*, Garth et al., *supra*

How, then, do attorneys' fees and gross settlement amounts compare in opt in FLSA collective actions? Of the 218 settled FLSA collective actions in my sample, data on both attorneys' fees and the gross settlement amount were available in fifty-seven cases. Figure 11 in the Appendix lists attorneys' fees, gross settlement amounts, and fee-settlement ratios for each of these cases showing that attorneys' fees represented 0% of the gross settlement at the low end and 97% at the high end, with a median of 50%. In other words, for every dollar the typical defendants paid in settling these FLSA collective actions, they paid fifty cents to the attorneys in fees. Figure 12 in the Appendix presents these figures in graphical form, showing attorneys' fees in relation to gross settlement amount. This 50% median fee-settlement ratio exceeds the 30%–32% figure from class action studies signaling that the opt in requirement may not, in fact, produce active plaintiffs able to police the trading of recoveries for fees.

Of course, there could be many explanations for the size of an attorney's fee, quite apart from any question of fee-maximization or fee-trading. The data do not reveal whether the median FLSA fee-settlement ratio exceeds the Rule 23(b)(3)

---

note 79, at 374 (labeling two types of private attorneys general the "legal mercenary" and the "social advocate"). This article does not distinguish among Rule 23(b)(3) class actions or FLSA collective actions on the basis of the type of entity that represented the plaintiffs. In any case, of the 250 FLSA collective actions examined in my data set, in only two were the plaintiffs represented by nonprofit organizations, Florida Legal Services, Inc. and the Partnership for Civil Justice. Given this extremely small sample, it would not be possible to draw conclusions about attorney behavior based on the nature of the entity representing the plaintiffs. This is perhaps an area for further research. It is interesting to note, however, that Professor Coffee predicts that the phenomenon of sweetheart deals "eventually will appear in those fields of law now dominated by the 'ideological' private attorney general," typified by the nonprofit "cause" lawyer. Coffee, *supra* note 3, at 236.

fee-settlement ratio for entirely legitimate reasons,[85] whether FLSA plaintiffs attempted to monitor attorneys' fee-maximizing behavior but failed, or whether the plaintiffs did not make the attempt at all. With respect to monitoring in particular—the focus of class action reformers—common sense suggests that the high costs to plaintiffs may dissuade them from making the attempt in the first place. Legal work can be opaque and hard to assess by a non-lawyer. The characteristics of FLSA plaintiffs outlined in Part VI.A, *supra*, including limited educational attainment and comprehension or language difficulties, may make monitoring even more difficult. In addition, the low wage workers who are FLSA plaintiffs may not have the leisure time or choose to spend the limited leisure time they have to monitor their attorneys. As a result, even if an especially motivated plaintiff were to try to monitor his or her attorneys, he or she would have to expend significant time, energy, and resources.

This problem of ensuring an attorney's fidelity, and its associated costs, is known in economic terms as an "agency cost."[86] As Susan Koniak and George Cohen explain in the Rule 23(b)(3) context:

> In ordinary lawyer-client relationships, clients can deal with the agency problem in two ways: by a contract between the lawyer and client that limits the lawyer's fee or ties it to the client's recovery, or by monitoring the lawyer's performance carefully as the representation proceeds. These solutions, imperfect enough in the ordinary client setting (especially when clients are unsophisticated individuals), are even more ineffective in the class action setting: 'The reason is that

---

[85] For example, the fee-shifting provision of the FLSA, 29 U.S.C. § 216(b), may enable plaintiffs' lawyers to demand higher fees in settlement, anticipating that the plaintiffs, as the prevailing party, would recover fees at the close of the litigation.

[86] Macey & Miller, *supra* note 5, at 12-13 (defining "the agency relationship as 'a contract under which one or more persons (the principal(s)) engage another person (the agent) to perform some service on their behalf which involves delegating some decision making authority to the agent.' In such cases . . . the interests of the agent are likely to deviate from those of the principal. This deviation of interests may not serve either principal or agent ex ante, and both, accordingly, may expend resources to overcome it.") (citation omitted).

> absent class members, by definition the majority of the class, neither contract with the lawyer, nor . . . monitor the lawyers' actions.' . . . Thus, client monitoring of lawyer performance is effectively unavailable in almost all class actions.[87]

The high fee-settlement ratios found in the FLSA collective actions in my study suggest that that agency costs of the sort identified by Koniak and Cohen are not limited to opt out Rule 23(b)(3) cases, but persist in FLSA cases despite the opt in requirement. In short, there is nothing to suggest that the mere act of opting in solves the agency problem or otherwise produces more effective, or less costly, monitoring.

In fact, the very existence of the FLSA's opt in requirement may discourage plaintiff monitoring, depress plaintiff recoveries, and drive up fee-settlement ratios. In a FLSA collective action, the plaintiffs are responsible for the cost of notice to potential opt in plaintiffs. In reality, this cost is borne by the plaintiffs' attorneys who seek reimbursement when the litigation ends. However, as the discussion of FLSA opt in rates in Part VI.A, *supra*, shows, even when notice issues, relatively few plaintiffs opt in. As a result, a rational plaintiffs' attorney might decide not to invest time and resources in the potentially futile notice process. This calculation is likely reflected in the fact that of the 250 cases in my FLSA data set that were captioned "collective action" when filed—signaling the existence of additional potential plaintiffs—the plaintiffs actually moved for collective action certification and permission to issue notice in only twenty-eight, or 11%.

Because notice is costly and generally ineffective at building the size of a case, plaintiffs' attorneys have a disincentive to pursue it. They have an incentive instead to adopt a high volume "portfolio" approach: to keep the number of plaintiffs small, to settle rapidly for a relatively low amount, to recover their investment of fees and costs, and to move on to the next case. Monitoring becomes more difficult if the number of plaintiffs in a case is relatively small, as each individual plaintiff must bear a greater share of the agency costs. In addition, plaintiff recoveries fall when attorneys pursue a portfolio strategy. The portfolio approach depresses plaintiff recoveries in two ways: by decreasing *total* plaintiff recovery by limiting the number of plaintiffs in the case, and by decreasing the *per-plaintiff* recovery in order to obtain a quick settlement.

---

[87] Koniak & Cohen, *supra* note 22, at 146 (quoting Susan P. Koniak & George M. Cohen, *Under Cloak of Settlement*, 82 VA. L. REV. 1051, 1104 (1996)).

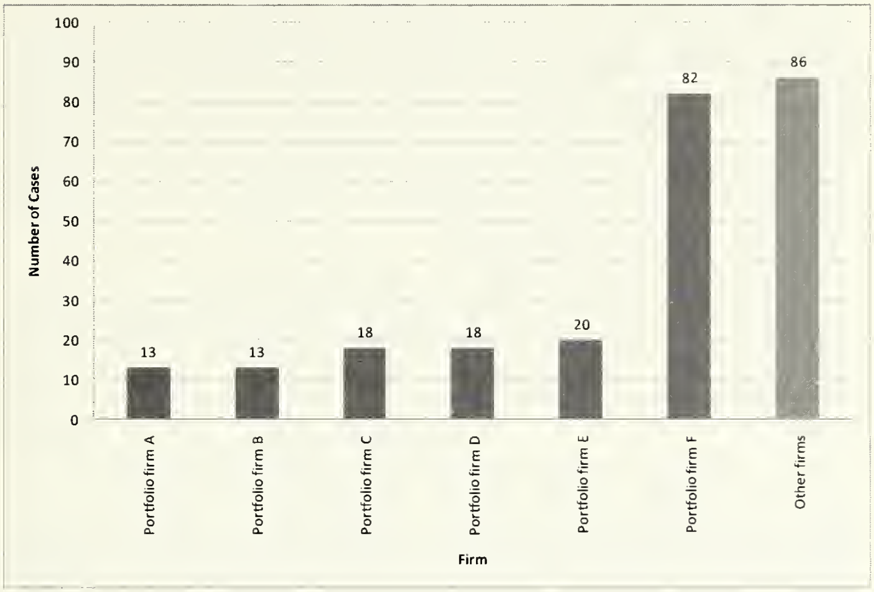
Not only will plaintiffs' recoveries be relatively low in these "portfolio" cases, but fee-settlement ratios will likely be relatively high. The fees negotiated by a portfolio attorney will represent, at minimum, the amount the attorney needs to keep the proverbial lights on. If the attorney is a fee-maximizer of the type identified by class action critics, then his or her fees will also include some additional premium. However, to the extent that all attorneys operating a for-profit law practice[88] have a fixed fee "floor" below which they cannot go, the lower the total settlement in any given case, the higher the fee-settlement ratio becomes. Put another way, if the attorneys' fees stay the same across all cases, the smaller the overall settlement "pie," the larger the slice representing attorneys' fees becomes, which then eats into the portion that remains for the plaintiffs. Thus, the cost and futility of the opt in process cause many attorneys to abandon it altogether resulting in ineffective plaintiff monitoring, relatively low plaintiff recoveries, and relatively high fee-settlement ratios.

Figure 8 suggests that plaintiffs' attorneys in my FLSA study have adopted this sort of portfolio approach. As Figure 8 shows, only six firms accounted for approximately 164 cases, or 66%, of the 250 cases in my data set. I call these six firms "portfolio firms."

---

[88] *See supra* note 84 and accompanying text.

Figure 8: FLSA Collective Actions Handled by Portfolio Firms and Other Firms



Data on attorneys' fees and gross settlement amounts were available for thirty-five cases handled by portfolio firms. Of the thirty-five cases, the plaintiffs' attorneys moved for collective action certification and permission to issue notice in only three cases, or 9%. However, all three cases settled before the end of the opt in period, meaning that additional potential plaintiffs did not have their full opportunity to join the case. The median plaintiffs' recovery in all thirty-five cases was $4,250.00; the median fee-settlement ratio was 50%.

Both fee and settlement data were available for twenty-two cases that were handled by other firms. In three of the twenty-two cases, the plaintiffs' attorneys moved for collective action certification and waited the entire opt in period before settling. The plaintiffs' attorneys moved for certification and then settled before the end of the opt in period in one additional case. In all, the plaintiffs' attorneys moved for certification and permission to issue notice in four of twenty-two cases, or 18%—double the rate in cases handled by portfolio firms. The median plaintiffs' recovery in all twenty-two cases was $6,425.00, and the median fee-settlement ratio was 49.7%.

The typical plaintiff represented by portfolio firms therefore recovered less than the typical plaintiff represented by other firms, and the fee-settlement ratio in a typical case handled by a portfolio firm was slightly higher than in a case handled by another firm. Moreover, plaintiffs' attorneys in non-portfolio firms moved for conditional collective action certification and requested permission to issue notice at twice the rate of attorneys in portfolio firms.

My study therefore suggests that instead of producing active client monitoring, the FLSA's opt in requirement may induce attorneys to forego notice, represent only a small number of plaintiffs, settle quickly for a relatively low amount, recover their fees, and move on to the next case in their portfolio. (Even in non-portfolio cases, most FLSA lawsuits that are captioned "collective action" never progress through the notice and opt in process and result in plaintiff recoveries and fee-settlement ratios that are only barely more plaintiff-friendly than in cases handled by portfolio firms.) In the class action context, this sort of attorney behavior is seen as the hallmark of the sweetheart deal.[89] However, the sweetheart deal narrative may not fit the FLSA context, as the FLSA's opt in requirement itself, rather than fee-maximizing behavior, might be contributing to the problem. Whether the driver of low plaintiff recoveries and high fee-settlement ratios is the opt in requirement, attorneys' fee-maximizing behavior, or some combination of both, the actual functioning of the FLSA's opt in requirement varies significantly from the expectations of class action reformers.[90]

---

[89] *See* Coffee, *supra* note 36, at 1366-67 (noting that the presence of "repeat players" can be a precondition for collusive, fee-maximizing behavior by attorneys).

[90] There is also evidence that, as in Rule 23(b)(3) class actions, there is little court monitoring of settlements and fee awards in FLSA collective actions. This is so even though the Eleventh Circuit, of which the Southern District of Florida is a part, has held that the parties must submit any FLSA settlement agreement to the court for a fairness determination. Lynn's Food Stores, Inc. v. United States, 679 F.2d 1350, 1353 (11th Cir. 1982) ("When employees bring a private action for back wages under the FLSA, and present to the district court a proposed settlement, the district court may enter a stipulated judgment after scrutinizing the settlement for fairness."). This requirement is designed to prevent plaintiffs from functionally waiving their FLSA rights

VII. CONCLUSION

In its title, this article asks whether an opt in requirement would fix the class action settlement. Put another way, are class action reformers correct in predicting that requiring plaintiffs to opt in would result in more meritorious claims filed, less attorney self-dealing at the settlement table, and ultimately greater plaintiff recoveries?[91] The answer suggested by my study of the opt in requirement of the Fair Labor Standards Act is a qualified "no." Recovery-loss ratios in settled FLSA collective actions are comparable to those in Rule 23(b)(3) class action settlements. The opt in requirement does not act as an accurate proxy for the merits of a case and therefore cannot be assumed to dictate attorneys' case filing decisions. Nor does the opt in requirement reduce fee-settlement ratios and may actually drive them up.

It would not be appropriate, however, on the basis of this study, to claim that plaintiffs' attorneys in FLSA cases in the Southern District of Florida are actually self-interested fee-maximizers who file strike suits and negotiate sweetheart

---

by agreeing to a rock-bottom settlement and to check the sort of sweetheart deals that worry class action critics. However, my review of FLSA case dockets has revealed that parties sometimes circumvent this requirement by filing a "stipulation of voluntary dismissal" with the court, without submitting their settlement agreement, making it impossible for the court to decide on fairness. *See, e.g.*, Agreed Motion for Approval of Settlement Stipulation and for Entry of Order of Dismissal with Prejudice at 2, Lipscomb v. Bochner, Inc., No. 07-20500 (S.D. Fla. May 31, 2007) (stating that "[a] confidential settlement agreement was negotiated on behalf of all of the parties, the terms of which reflect a reasonable compromise of the parties' many disputed issues" and that the agreement satisfied the requirements of *Lynn's Food Stores*); Joint Stipulation of Voluntary Dismissal with Prejudice at 1, Alfasi v. Lydian Data Services, LLC, No. 05-60758 (S.D. Fla. Sept. 1, 2005) (stating only that "[t]he parties have reached an amicable resolution of this matter."). The judges in these cases did not appear to object to this practice. When the parties do submit a proposed settlement agreement, it is more often than not approved without question by the court, leaving the parties' agreement to stand. *But see* Silva v. Miller, 547 F. Supp. 2d 1299 (S.D. Fla. 2008) (engaging in a searching review of the parties' settlement agreement and the plaintiffs' attorneys' fees).

[91] In his seminal article on public law litigation, Professor Abram Chayes asked the same question in a different way: "And in the absence of a particular client, capable of concretely defining his own interest, can we rely on the assumptions of the adversary system as a guide to the conduct and duty of the lawyer?" Abram Chayes, *The Role of the Judge in Public Law Litigation*, 89 HARV. L. REV. 1281, 1291 (1976).

deals. While the data on recovery-loss ratios, opt in rates, and fee-settlement ratios may appear at first glance to support such labels, a closer look reveals a more complicated picture with the opt in requirement itself likely playing a role in keeping plaintiffs' recoveries and participation rates low and driving up attorneys' fees in relation to plaintiffs' recoveries.

The story of the FLSA's opt in requirement is more appropriately read as a cautionary tale, signaling the unintended consequences of transforming Rule 23(b)(3) into an opt in regime. Not only does an opt in requirement appear *not* to accomplish the goals of class action reformers, but it may cause real harm to plaintiffs by creating serious disincentives for them to pursue even meritorious claims and contributing to low total plaintiff recoveries in the cases that are filed. In many ways, the opt in requirement of the FLSA seems to eliminate any possibility of collective action, making the FLSA enforceable only by individuals or very small groups.[92] The opt in requirement may, therefore, undermine the project of law enforcement generally, dissuading plaintiffs from acting collectively as private attorneys general to enforce the FLSA's minimum wage and overtime mandates.[93]

Further research remains to determine the extent to which these findings are unique to the FLSA. Studies of opt in Equal Pay Act cases, Age Discrimination in Employment Act cases, and state law opt in class actions would provide additional windows into the workings of the opt in requirement in other opt in regimes. Additional study is also necessary to investigate how well the FLSA would predict the workings of an opt in requirement in the variety of cases that are currently filed as Rule 23(b)(3) class actions. Would an opt in requirement function in Title VII, securities, and mass tort class actions as in the FLSA, for example? My study suggests that, whatever the claim type, plaintiffs who are in an ongoing relationship with the defendants whom they are suing would likely feel the same pressures against participation and fear of retaliation that face potential FLSA opt in plaintiffs. In the Rule 23(b)(3) context, such plaintiffs might include those in Title VII or other employment class actions.[94] Likewise, plaintiffs, who are particularly economically vulnerable, or those who have suffered an accident or illness in a Rule 23(b)(3) mass injury class action, might be physically or financially unable to expend the time and energy required to opt in. Finally, plaintiffs who would be vulnerable to public scrutiny or opprobrium if they joined a lawsuit—one might imagine plaintiffs suffering from a socially stigmatized medical condition whose health status is at issue in a case, for example—might be unwilling to make the public self-identification required by the opt in requirement.[95] By contrast, the relatively more powerful and secure plaintiffs in Rule 23(b)(3) securities class actions might not be subject to the same vulnerabilities as FLSA plaintiffs, and conclusions drawn from the FLSA might not be as applicable.[96]

---

[92] Whether this result is good or bad depends on one's party alignment. Defendants, for example, likely would not lament the elimination of a collective action option in many statutory regimes. *See supra* note 58.

[93] For a discussion of the FLSA as a statute designed to be enforced by private attorneys general, see Brooklyn Sav. Bank v. O'Neil, 324 U.S. 697, 709 (1945). For a discussion of the compensation and deterrence goals of private attorney general class actions, see Coffee, *supra* note 3, at 220 and Stephen Berry, *Ending Substance's Indenture to Procedure: The Imperative for Comprehensive Revision of the Class Damage Action*, 80 COLUM. L. REV. 299, 300 (1980).

---

[94] *Developments in the Law: Class Actions*, 89 HARV. L. REV. 1318, 1443 n.262 (1976) ("In cases in which the defendant has some implicit leverage over class members – such as an employer would have over employee class members in a Title VII action – requiring the absentee to opt in can effectively decimate the class."). Another relevant future research project could be a comparison of opt in rates in FLSA, EPA, and ADEA cases (all employment litigation) by plaintiffs who remain employed by the defendant and those who are former employees. This would help isolate the variables, including an ongoing relationship between the parties that may suppress plaintiff participation in an opt in regime. In addition, a comparative study of group litigation in countries where opt in is already the default rule might be fruitful. *See, e.g.*, Antonio Gidi, *Class Actions in Brazil—A Model for Civil Law Countries*, 51 AM. J. COMP. L. 311, 338 n.63 (2003) (discussing opt in regimes in France, Japan, and Scotland); Deborah R. Hensler, *The Globalization of Class Actions: An Overview*, 622 ANNALS 7, 16 (2009); Coffee, *supra* note 15, at 57-58.

[95] *But cf.* Robert L. Carter, *The Federal Rules of Civil Procedure as a Vindicator of Civil Rights*, 137 U. PA. L. REV. 2179, 2186 (1989) (noting that in desegregation cases, "[a] lone plaintiff was extremely vulnerable to the pressure of intimidation by state and local officials, and it was not above those officials to bring such pressure to bear").

[96] *See also supra* note 72.

One possible area for additional inquiry into these questions would be to compare the rate at which plaintiffs in various types of Rule 23(b)(3) class actions file damages claim forms at the end of a case. Such claim-filing procedures function essentially as mini-opt in requirements within the larger Rule 23(b)(3) opt out regime. Research suggests that claim-filing rates in Rule 23(b)(3) consumer, mass tort, and securities class actions are typically low.[97] However, distinguishing among these claim types—more accurately, among plaintiff types—might reveal interesting differences in claim filing rates, suggesting a possible differential impact across claim-types and plaintiff-types if an opt in requirement were enacted in the context of Rule 23(b)(3). This may, in turn, suggest a differently-structured opt in reform to Rule 23(b)(3) class actions depending on the claim type or plaintiff type at issue. As an anti-retaliation protection, for example, plaintiffs in employment cases who would be required to sue their employers could be permitted to use a pseudonym on their opt in form, with their identity available only to the court and to counsel.

The analyses presented in this article of recovery-loss ratios, opt in rates, and fee-settlement ratios in FLSA collective actions support the preliminary conclusion that giving a greater role to the absent plaintiff class, through an opt in requirement, would not accomplish Rule 23(b)(3) class action settlement reform. In short, the "plaintiff" lever may be the wrong one to push. Potential plaintiffs simply do not appear to make opt in decisions based solely on the merits or to be able to monitor their attorneys effectively once they have joined a FLSA collective action. If class action critics seek to regulate what they see as self-interested, fee-maximizing attorney behavior in the form of strike suits and sweetheart deals, reforms that taget those behaviors directly would likely be both more appropriate and more effective.[98]

---

[97] HENSLER ET AL., *supra* note 17, at 427 (examining consumer and mass tort class actions); Coffee, *supra* note 15, at 21; Issacharoff & Miller, *supra* note 35, at 205 (discussing securities class actions); Cox & Thomas, *supra* note 35, at 412 (same).

---

[98] Possible reforms include more searching judicial oversight of settlements, a requirement that fees be calculated as a percentage of the plaintiffs' recovery, and decoupling the negotiation of fees from the negotiation of the plaintiffs' recovery.

APPENDIX

Figure 9: Plaintiffs' Recoveries, Losses, and Recovery-Loss Ratios in Settled FLSA Collective Actions (Sorted by Plaintiffs' Loss)

Case Number[99] (N=25)	Case	Plaintiffs' Recovery	Plaintiffs' Loss	Recovery-Loss Ratio
1:07-cv-21476-CMA	Longas v. NIDO LLC et al	$2,063.00	$2,063.00	100%
1:07-cv-23082-MGC	Cuadra v. BLU System Corporation et al	$3,000.00	$4,389.00	68%
1:08-cv-21171-PCH	Taveras v. Bayus Security Services, Inc. et al	$3,100.00	$7,552.00	41%
0:08-cv-60945-JIC	Delgardio v. Foremost Plumbing, Inc. et al	$2,500.00	$7,680.00	33%
9:08-cv-81375-WJZ	Dieujuste v. Quality Sod of the Palm Beaches, Inc. et al	$4,000.00	$7,800.00	51%
1:06-cv-22971-JLK	Recio et al v. Altex, Inc. et al	$9,000.00	$10,065.00	89%
0:07-cv-61777-JAL	Metcalf v. iBill, LLC	$6,500.00	$10,946.89	59%
1:08-cv-21869-DLG	Garcia v. MG Bus Service & Tours, Inc. et al	$450.00	$12,300.00	4%
0:08-cv-60102-CMA	Rodrigues v. AF of Fort Lauderdale, Corp et al	$2,437.50	$13,096.80	19%
2:07-cv-14334-KMM	Bianchini v. El Dorado Financial, Inc. et al	$1,000.00	$21,600.00	5%
0:08-cv-61728-JJO	Borges v. The Keiser School, Inc.	$4,500.00	$23,165.68	19%
0:06-cv-60021-DLG	Williams v. Modern Concrete, LLC	$30,330.70	$31,200.00	97%
0:07-cv-61690-WJZ	Cohen v. Allied Steel Buildings, Inc. et al	$32,000.00	$36,152.23	89%
1:08-cv-20019-JEM	Ayala v. Capform, Inc. et al	$6,000.00	$38,880.00	15%
0:08-cv-61792-FAM	Isaacs v. Gold Coast Tire Holdings, Inc. et al	$9,000.00	$42,400.00	21%

---

[99] *Johnson* does not appear in Figure 4, *supra*, because the recovery and loss figures themselves were not available, but the parties' settlement agreement states that the plaintiffs recovered "approximately 39% of the full amount of their wages and liquidated damages." Plaintiffs' Unopposed Verified Motion to Approve Settlement Agreement, and Memorandum of Law at 9, Johnson v. Senior Health Care Servs., No. 05-60871-MGC (S.D.F.L. May 31, 2006), Docket entry 209.

1:07-cv-22561-ASG	Torcios v. Esther's 103 Inc. et al	$3,800.00	$46,800.00	8%
1:06-cv-20884-PAS	Auguste v. Choice Hotels Int'l, et al	$7,854.00	$51,660.00	15%
9:06-cv-80791-KAM	Pierre v. Shiv Shakti Donut, et al	$6,350.00	$56,160.00	11%
2:08-cv-14150-DLG	Cadet v. Pedersen Lathing & Plastering, Inc. et al	$23,000.00	$56,940.00	40%
0:07-cv-60540-WJZ	Prescia et al v. Mancini Ventures, LLC et al	$3,690.00	$63,395.00	6%
1:07-cv-22232-WMH	Medeiros v. Kanaan Racing,Inc et al	$29,500.00	$70,092.66	42%
1:07-cv-22068-FAM	Ore v. U.S. Security Associates, Inc.	$29,554.24	$123,060.00	24%
1:04-cv-21861-ASG	Reyes, et al v. Carnival Corporation	$60,500.00	$217,500.00	28%
0:05-cv-60099-WPD	Plaseyed, et al v. PCA National,LLC	$39,050.07	$244,062.93	16%
0:05-cv-60871-MGC	Johnson, et al v. Senior Health Care, et al			39%
			Median:	28%

Figure 10: Notices, Opt Ins, and Opt In Rates in Complete FLSA Collective Actions (Sorted by Number of Notices)

Case Number (N=38)	**Case**	**Opt In Period** (days)[100]	**Form of Notice**	**Notic tices**[101]	**Opt Ins**	**Opt In Rate**
0:05-cv-60292-DLG	Spencer, et al v. Regional Acceptance, et al		Mail	3	1	33%
0:03-cv-61234-WJZ	Roseberry, et al v. Answer Group, Inc.	30	Mail	4	0	0%
1:06-cv-22055-JLK	Flores, et al v. Park West Parking, et al	60	Mail	7	1	15%
9:03-cv-80908-WPD	Belfleur, et al v. Mignano Lawn Care, et al	90	Mail	13	2	15%

---

[100] The length of the opt in period is listed when this information was available in the court's conditional collective action certification order.

[101] The opt in rates in shaded rows are estimates provided by the plaintiffs' attorneys. In these cases, the case files were in storage or otherwise inaccessible, so the attorneys could not provide the exact number of notices that were sent to potential opt in plaintiffs.

Case Number (N=38)	Case	Opt In Period (days)[100]	Form of Notice	Notictices[101]	Opt Ins	Opt In Rate
0:07-cv-61010-WJZ	Godoy v. New River Pizza, Inc. et al		Mail	13	2	15%
9:06-cv-80505-JCP	Reese v. Kimberly Credit Coun. et al		Mail	13	2	15%
9:06-cv-80380-KLR	Hernandez, et al v. TFC USA, Corp., et al	60	Mail	20	3	15%
9:07-cv-80240-DTKH	Patterson v. Palm Beach County School Board	60	Mail	27	4	15%
9:06-cv-80551-DMM	Riffin, et al v. Rapcon Incorporated, et al	25	Mail	33	5	15%
9:03-cv-80378-JCP	Charles, et al v. Prime Management Gro		Mail	40	6	15%
9:04-cv-80178-KAM	Norzelus, et al v. Republic Services	90	Mail	40	6	15%
9:07-cv-80272-KAM	Dieujuste v. R. J. Electric, Inc. et al	60	Mail	47	7	15%
9:06-cv-80707-WPD	St. Louis, et al v. Express Car Wash	30	Mail	47	7	15%
0:03-cv-61534-CMA	Bergeron, et al v. Watson Trucking, et al	60	Mail	48	7	15%
0:08-cv-61378-ASG	McKenzie v. Lindstrom Air Conditioning, Inc. et al		Mail	53	8	15%
9:06-cv-80604-DMM	Butler, et al v. Oak Street Mortgage	60	Mail	68	16	24%
9:07-cv-81013-KAM	Vaccaro v. Candidates On Demand Group, Inc.	60	Mail	73	11	15%
1:05-cv-20310-JEM	Luna, et al v. Hereford Grill	30	Mail	75	0	0%
0:05-cv-61757-WPD	Morrison, et al v. Quality Transport	60	Mail	75	16	21%
0:06-cv-61299-PAS	Beaudouin v. S.W. Bach & Company		Mail	80	12	15%
9:08-cv-80004-KLR	Russo v. Cold Air Distributors Warehouse of Florida, Inc.		Mail	87	13	15%
0:03-cv-61177-AMS	Knouse, et al v. Florida Power	120	Mail	100	19	19%

Case Number (N=38)	Case	Opt In Period (days)[100]	Form of Notice	Notic tices[101]	Opt Ins	Opt In Rate
2:05-cv-14295-KAM	Carmody, et al v. Florida Center	45	Mail	106	20	19%
1:05-cv-21887-TK	Blanco, et al v. Federal Detective, et al	60	Mail	140	3	2%
9:06-cv-80645-DTKH	Rusignuolo v. McLane Suneast, Inc., et al	30	Mail	160	24	15%
0:05-cv-60871-MGC	Johnson, et al v. Senior Health Care, et al		Mail	161	77	48%
0:03-cv-62091-WPD	Barrett, et al v. Capital Acquisitions	90	Mail	200	18	9%
0:07-cv-61391-WPD	Petit et al v. ABC Cutting Contractors, Inc.		Mail	207	31	15%
1:07-cv-21824-DLG	King et al v. CVS/Caremark Corp. et al	45	Mail	225	17	8%
0:07-cv-61134-WPD	Rivera v. Gennaro's Produce, Inc. et al	90	Mail	246	19	8%
2:07-cv-14156-FJL	Arena et al v. Mortgage U.S., Inc. et al	30	Mail	249	67	27%
0:05-cv-61865-WPD	Adams, et al v. ABN Amro, Inc., et al	30	Mail	339	149	44%
0:08-cv-61254-UU	Titre v. Essia et al	105	Mail & posting at worksite	353	53	15%
1:07-cv-22000-UU	Walters et al v. American Coach Lines of Miami, Inc.	60	Mail	393	60	15%
1:04-cv-21861-ASG	Reyes, et al v. Carnival Corporation	120	Mail	580	87	15%
2:05-cv-14059-KMM	Ealy-Simon, et al v. Liberty Medical, et al	30	Mail	4,229	311	7%
9:04-cv-80521-KAM	Pendlebury, et al v. Starbucks Coffee, et al	60	Mail	7,500	900	12%
0:05-cv-60099-WPD	Plaseyed, et al v. PCA National,LLC	30	Mail	25,499	1,560	6%
				Median:		15%

Figure 11: Attorneys' Fees, Gross Settlement Amounts, and Fee-Settlement Ratios in Settled FLSA Collective Actions (Sorted by Gross Settlement Amount)

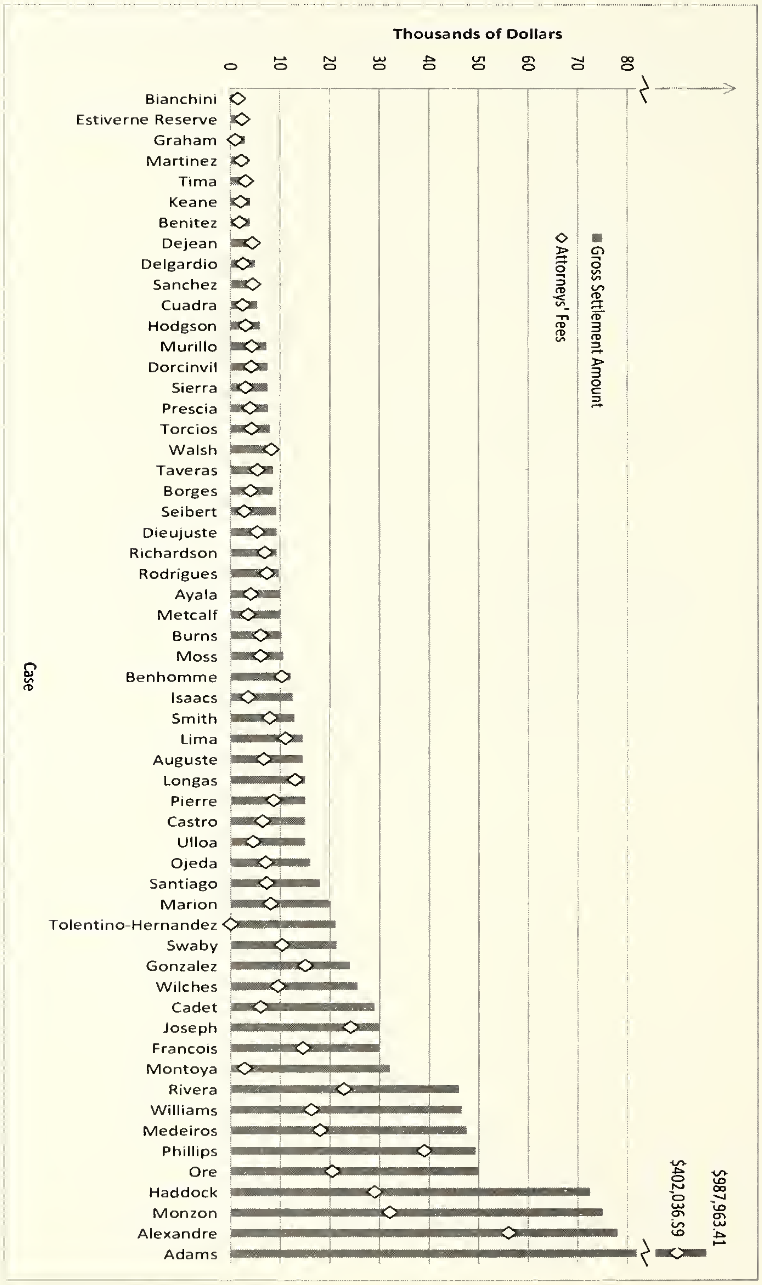
Case Number (N=57)	Case	Attorneys' Fees	Gross Settlement Amount	Fee-Settlement Ratio
2:07-cv-14334-KMM	Bianchini v. El Dorado Financial, Inc. et al	$1,500.00	$2,500.00	60%
9:07-cv-80250-DTKH	Estiverne Reserve v. Federal Express Corporation	$2,300.00	$3,000.00	77%
9:07-cv-80935-DTKH	Graham v. Citi Trends, Inc.	$1,000.00	$3,000.00	33%
9:08-cv-80447-KLR	Martinez v. Wellington Granite & Marble,Inc. et al	$2,235.50	$3,500.50	64%
9:07-cv-80689-WPD	Tima v. Premier Beveral Company, LLC.	$3,000.00	$4,000.00	75%
0:06-cv-61408-DTKH	Keane v. Knights Towing, Inc.	$2,000.00	$4,000.00	50%
9:08-cv-80277-KLR	Benitez et al v. Mahood Cleaning Services,Inc. et al	$1,829.52	$4,000.02	46%
0:06-cv-60895-WPD	Dejean v. Jay Krishna Enterpri, et al	$4,400.00	$5,000.00	88%
0:08-cv-60945-JIC	Delgardio v. Foremost Plumbing, Inc. et al	$2,500.00	$5,000.00	50%
1:03-cv-21789-PCH	Sanchez v. South Beach Group	$4,450.00	$5,250.00	85%
1:07-cv-23082-MGC	Cuadra v. BLU System Corporation et al	$2,400.00	$5,400.00	44%
1:06-cv-22465-MGC	Hodgson v. Corporate Protection, et al	$3,000.00	$6,000.00	50%
1:05-cv-21763-BLG	Murillo v. Intercontinental H	$4,250.00	$7,250.00	59%
9:07-cv-80922-DMM	Dorcinvil v. Mar-A-Lago Club, LLC et al	$4,175.00	$7,500.00	56%
9:08-cv-	Sierra v. The Pep Boys-	$3,000.00	$7,500.00	40%

Case Number (N=57)	Case	Attorneys' Fees	Gross Settlement Amount	Fee-Settlement Ratio
81580-KLR	Manny, Moe & Jack Inc			
0:07-cv-60540-WJZ	Prescia et al v. Mancini Ventures, LLC et al	$3,910.00	$7,600.00	51%
1:07-cv-22561-ASG	Torcios v. Esther's 103 Inc. et al	$4,200.00	$8,000.00	53%
9:03-cv-81032-JCP	Walsh v. Auction Warehouse, et al	$8,200.00	$8,494.38	97%
1:08-cv-21171-PCH	Taveras v. Bayus Security Services, Inc. et al	$5,400.00	$8,500.00	64%
0:08-cv-61728-JJO	Borges v. The Keiser School, Inc.	$4,000.00	$8,500.00	47%
9:05-cv-81153-DTKH	Seibert v. Harbor Title, L.C.	$2,750.00	$9,250.00	30%
9:08-cv-81375-WJZ	Dieujuste v. Quality Sod of the Palm Beaches, Inc.	$5,350.00	$9,350.00	57%
1:07-cv-21289-KMM	Richardson v. TB Isle Resort LP, et al	$6,875.00	$9,353.80	73%
0:08-cv-60102-CMA	Rodrigues v. AF of Fort Lauderdale, Corp et al	$7,312.50	$9,750.00	75%
1:08-cv-20019-JEM	Ayala v. Capform, Inc. et al	$4,000.00	$10,000.00	40%
0:07-cv-61777-JAL	Metcalf v. iBill, LLC	$3,500.00	$10,000.00	35%
0:06-cv-61809-JIC	Burns v. Carevacations Carribean Ports Medical Inc.,	$6,000.00	$10,250.00	59%
0:08-cv-61239-UU	Moss v. Baby Borrow Rental Corp. et al	$6,000.00	$10,600.00	57%
9:03-cv-80577-JCP	Benhomme v. Kids To Kidz, Inc, et al	$10,275.00	$12,134.97	85%
0:08-cv-61792-FAM	Isaacs v. Gold Coast Tire Holdings, Inc. et al	$3,500.00	$12,500.00	28%
9:05-cv-80440-	Smith, et al v. Mobile Transactions, et al	$7,815.00	$12,815.00	61%

Case Number (N=57)	Case	Attorneys' Fees	Gross Settlement Amount	Fee-Settlement Ratio
JCP				
9:07-cv-81024-DTKH	Lima v. Hometown Pest Control, Inc. et al	$11,000.00	$14,500.00	76%
1:06-cv-20884-PAS	Auguste v. Choice Hotels Int'l, et al	$6,646.00	$14,500.00	46%
1:07-cv-21476-CMA	Longas v. NIDO LLC et al	$12,937.00	$15,000.00	86%
9:06-cv-80791-KAM	Pierre v. Shiv Shakti Donut, et al	$8,650.00	$15,000.00	58%
0:07-cv-61561-MGC	Castro v. Gold Coast Beverage Distributors, Inc.	$6,400.00	$15,000.00	43%
0:07-cv-61486-UU	Ulloa v. National Art Landscape & Design, Inc.	$4,500.00	$15,000.00	30%
9:08-cv-80558-KLR	Ojeda v. WHM, LLC	$7,000.00	$16,000.00	44%
1:07-cv-21513-WMH	Santiago v. Elbar, Inc. et al	$7,200.00	$18,000.00	40%
2:07-cv-14124-DLG	Marion v. Prestige Tax, Inc. et al	$8,000.00	$20,000.00	40%
2:07-cv-14382-KMM	Tolentino-Hernandez et al v. Becerra	$0.00	$21,148.05	0%
0:06-cv-61401-PAS	Swaby v. Carevacations Caribb	$10,350.00	$21,350.00	48%
1:06-cv-20260-CMM	Gonzalez v. Qualender Food Corp., et al	$15,000.00	$24,000.00	63%
9:06-cv-80306-JCP	Wilches v. Streamline Construct, et al	$9,500.00	$25,500.00	37%
2:08-cv-14150-DLG	Cadet v. Pedersen Lathing & Plastering, Inc. et al	$6,000.00	$29,000.00	21%
9:07-cv-80032-DTKH	Joseph v. Boca Raton Community Hospital Inc.	$24,100.00	$30,000.00	80%
0:06-cv-61541-UU	Francois v. Atlantic Roofing of Florida II, Inc., et al.	$14,513.00	$30,000.00	48%

Case Number (N=57)	Case	Attorneys' Fees	Gross Settlement Amount	Fee-Settlement Ratio
9:08-cv-80883-DTKH	Montoya v. Physicians Surgical Group, L.L.C. et al	$2,800.00	$32,050.00	9%
0:07-cv-61134-WPD	Rivera v. Gennaro's Produce, Inc. et al	$22,800.10	$46,000.00	50%
0:06-cv-60021-DLG	Williams v. Modern Concrete, LLC	$16,225.90	$46,556.60	35%
1:07-cv-22232-WMH	Medeiros v. Kanaan Racing,Inc et al	$18,000.00	$47,500.00	38%
9:07-cv-80140-DTKH	Phillips v. Palm Beach County Fire Rescue Dept.	$39,000.00	$49,373.75	79%
1:07-cv-22068-FAM	Ore v. U.S. Security Associates, Inc.	$20,445.76	$50,000.00	41%
0:06-cv-61505-WJZ	Haddock, et al v. Holland America Line, et al	$29,000.00	$72,500.00	40%
1:03-cv-20124-JAL	Monzon, et al v. Amaralto Concrete	$32,028.51	$75,000.00	43%
0:07-cv-60099-WPD	Alexandre v. Gold Coast Beverage Distributors, Inc.	$56,000.00	$78,000.00	72%
0:05-cv-61865-WPD	Adams, et al v. ABN Amro, Inc., et al	$402,036.59	$1,390,000.00	29%
			Median:	50%

Figure 12: Plaintiffs' Recoveries and Attorneys' Fees in Settled FLSA Collective Actions (Sorted by Plaintiffs' Recovery)



FORM AND SUBSTANCE OF THE DOCTRINE OF PIERCING THE CORPORATE VEIL

Thomas K. Cheng[*]

INTRODUCTION 498

I. THE FORMAL RULE-STANDARD DICHOTOMY OF THE CORPORATE VEIL DOCTRINE 501

- *A. Limited Liability as a Rule and the Corporate Veil Doctrine as a Standard* 501
- *B. The Practical Implications of the Formal Rule-Standard Dichotomy* 510
- *C. The Corporate Veil Doctrine as a Hybrid Formalities-Conduct Directive* 515

II. THE SUBSTANTIVE INDIVIDUALISM-ALTRUISM DICHOTOMY OF THE CORPORATE VEIL DOCTRINE 521

- *A. Limited Liability as an Individualistic Rule and the Corporate Veil Doctrine as an Altruistic Standard* 521
- *B. Objectives of the Corporate Veil Doctrine* 525
  1. Punishment for Improper Conduct 528
  2. Compensation for the Corporate Creditors 530
  3. Deterrence of Future Improper Conduct 534
  4. Prevention of Unjust Enrichment of the Shareholders 540
  5. Summing Up 547

III. REFORMULATION OF THE CORPORATE VEIL DOCTRINE BASED ON AN UNJUST ENRICHMENT RATIONALE 550

- *A. The Formalities Prong* 551
  1. Non-observance of Corporate Formalities 551
  2. Lack of Substantive Separation 556
  3. Shareholder Domination and Overlap in Corporate Personnel 559
- *B. The Impropriety Prong* 565

---

[*] The author is an assistant professor in the Faculty of Law of the University of Hong Kong. He received his B.A. from Yale College, his J.D. from Harvard Law School, and his B.C.L. degree (equivalent of LL.M.) from University of Oxford.

1. Inadequate Capitalization 567
    - *(i) Whether Inadequate Capitalization Should be a Relevant Factor*..569
    - *(ii) The Appropriate Test for Capital Adequacy*....572
    - *(iii) Different Analyses for Contract and Tort Cases*...578
    - *(iv) Summing Up*..580
2. Other Improper Conduct..581

C. Remedy...582

CONCLUSION 585

INTRODUCTION

The doctrine of piercing of the corporate veil remains one of the most confused areas of corporation law.[1] This is so despite the fact that, together with the related topic of limited shareholder liability, it is one of the most extensively written and closely examined topics of corporation law. The courts have not gone much beyond circumstance-specific inquiries into the equities of each case. Commentators have struggled to distill clear trends in or come up with an overarching theory that accounts for the case law. Given the volume and variety of the cases, the search for such a theory may remain elusive. This is not to say, however, that a more focused approach to veil piercing cannot be formulated upon a clearer understanding of the doctrine.

Much of the academic literature on the topic has either been doctrinal analysis of the case law or theoretical examination of the economic basis of limited liability. There is a distinct lack of theoretical analysis of the doctrine itself. For instance, the existing literature sheds little light on the objective of the doctrine. What is the purpose served by the doctrine? Does it merely reinforce the functions of the area of law that underlies the corporate veil claim? Or does it serve an independent purpose? If it does, is it compensation, deterrence, punishment, or something else altogether? There is also little systematic evaluation of the relationship between the doctrine and limited liability, the general rule to which the doctrine serves as an exception. The purpose of this Article is to examine the formal and the substantive relationships between the doctrine and limited liability and to ascertain the objective of the corporate veil doctrine. With the insights gleaned from this exercise, the doctrine will be reoriented so that it focuses on what truly motivates the courts to pierce the veil.

---

[1] This article will only consider cases in which the corporate veil is pierced to impose liability on shareholders. There are a considerable number of cases in which the veil is pierced for other purposes, such as contractual construction or establishment of jurisdiction. *See* Robert W. Hamilton, *The Corporate Entity*, 49 TEX. L. REV. 979, 1002-07 (1971). However, this article will not focus on them. Only in those veil piercing cases in which liability is imposed on the shareholders is the principle of limited liability at stake.

Applying the formal dichotomy of rules and standards and the substantive dichotomy of individualism and altruism developed by Professor Kennedy in his seminal article "*Form and Substance in Private Law Adjudication*," this Article ascertains the formal and substantive relationships between limited liability and the corporate veil doctrine.[2] It will be shown that in formal terms, limited liability shows all the relevant characteristics of a legal rule while the corporate veil doctrine functions as a standard. In substantive terms, limited liability is distinctly individualistic in nature while the corporate veil doctrine possesses an altruistic character. Analysis along these two formal and substantive dimensions enhances our understanding of the corporate veil doctrine. In particular, the altruistic character of the corporate veil doctrine will be used as a basis upon which to launch an inquiry into the objective of the doctrine. After an assessment of the effects of veil piercing and its interaction with the underlying contractual or tortious liability, it will be shown that the principal objective of the doctrine is in fact the prevention of the unjust enrichment of the shareholders at the expense of the corporation's creditors. What ultimately drives the court's decision to pierce the veil is the fact that the shareholders have benefited to the detriment of the creditors in an unjust manner.

---

[2] *See Id.* Corporation law technically is not an area of private law. However, many of the insights developed by Professor Kennedy are applicable to areas outside of private law. At least from a formal perspective, there is nothing distinctive about rules and standards in private law as opposed to their counterparts in public law and other realms of law. This is consistent with the fact that Kennedy's discussion about the formal characteristics of rules and standards draws on examples from other areas of law such as criminal law and constitutional law.

This Article proposes a reconsideration of the doctrine in light of this shareholder enrichment rationale. In particular, it proposes that veil piercing analysis should focus on conduct that results in shareholder enrichment at the expense of the creditors. Other considerations either only reflect the probability of shareholder enrichment, in which case it would be better to focus on the fact of enrichment itself, or worse still obscure the analysis. Applying this prevention of shareholder enrichment rationale to the two main components of the doctrine, the formalities prong and the impropriety prong, this Article suggests that most of the criteria under the first prong, such as observance of corporate formalities, shareholder domination, overlap in corporate personnel, and lack of substantive separation between the shareholders and the corporation, do not focus on the most relevant consideration in the corporate veil cases. Instead, the courts should draw their attention to instances of inadequate capitalization, asset stripping, deliberate unprofitable operation of a corporation, and other similar conduct that enriches the shareholders at the expense of outside creditors.

It is important to note that the discussion in this Article will be confined to contractual and tortious corporate veil cases. There are a number of reasons for this. First, the statutory corporate veil cases present different policy issues and require a different mode of analysis from the contractual and tortious cases. Determination of the statutory cases must take into account the policy of the underlying statutes and often hinges on statutory interpretation. For example, veil piercing under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) may depend on the court's interpretation of the term "operator" and how the court reconciles limited liability with the environmental policy underpinning CERCLA.[3] Veil piercing in taxation cases requires a balancing of "the broad social and governmental policies under federal income tax laws" and the policy justifications for limited liability.[4] Second, confining our discussion to the corporate veil cases premised on private law claims is appropriate in light of this Article's invocation of Professor Kennedy's formal and substantive dichotomies, which were developed in the context of private law adjudication. The individualism-altruism distinction has limited application to the statutory cases.

---

[3] United States v. Bestfoods, 524 U.S. 51, 67-68 (1998).

This Article is divided into four sections. Section I examines the formal relationship between limited liability and the corporate veil doctrine as a rule and a standard respectively. Both theoretical and practical implications for the corporate veil doctrine will be drawn from this examination. Section II delves into the substantive relationship between limited liability and the corporate veil doctrine and confirms their respective individualistic and altruistic nature. Building on this altruistic foundation, an inquiry will be made into the principal objective of the doctrine. It will be shown that the doctrine is ultimately concerned with the prevention of the unjust enrichment of shareholders at the expense of creditors. Section III proposes a reformulation of the doctrine in light of this newly ascertained rationale. In particular, it will be argued that the various criteria under the formalities prong of the doctrine should be de-emphasized, as they fail to focus on the crux of the veil piercing cases. Section IV concludes the Article.

I. THE FORMAL RULE-STANDARD DICHOTOMY OF THE CORPORATE VEIL DOCTRINE

A. Limited Liability as a Rule and the Corporate Veil Doctrine as a Standard

The default rule in corporation law is that shareholders enjoy limited liability protection. Their exposure to corporate liabilities is confined to their equity investments in the firm. They can lose no more than the value of their entire shareholding in the event of a corporate failure. A related rule to limited liability is separate corporate personality—the notion that a corporation is an entity distinct from its shareholders in the eyes of the law. The two rules are closely related, but are not synonymous. While holding shareholders liable for corporate debts entails ignoring the separate corporate personality, the corporate entity can be set aside without resulting in shareholder liability.[5] Recognition of separate corporate personality means that the corporation possesses the capacity to enter into contracts and own property in its own name, and the capacity to sue and be sued. It is capable of committing torts, for which it alone is liable except under narrow circumstances under which others may be held responsible.[6] Moreover, conduct of the corporation is generally not attributable to its shareholders and vice versa. These are simple and clear-cut rules that are of general applicability. As long as a corporation is duly incorporated, it functions as a distinct legal entity, and its shareholders enjoy limited liability protection.

---

[4] Hamilton, *supra* note 1, at 998.

The piercing of corporate veil doctrine is an exception to the general rules of limited liability and separate corporate personality.[7] Under the doctrine, limited liability protection for shareholders and separate corporate personality may be overridden if certain conditions are met. The courts have repeatedly asserted that the doctrine is an equitable one and requires a weighing of the totality of the circumstances.[8] There are a number of formulations of the doctrine. Many states have promulgated a two-prong test, requiring a showing that first, there is such a unity of interest and ownership between the corporation and its owners that their separate personalities have ceased to exist in reality, and second, an adherence to separate corporate personality would create inequitable results.[9] A commentator has called the first prong the formalities prong and the second the unfairness prong.[10] In his authoritative survey of the doctrine, Professor Thompson identified the following substantive factors as being most frequently considered by the courts in their analysis: lack of substantive separation between the shareholders and the corporation, intertwining, inadequate capitalization of the corporation, non-observance of corporate formalities, shareholder domination, and overlap of corporate personnel and management.[11] These factors are mostly relevant to the question of unity of interest and ownership. Another authoritative formulation is the instrumentality doctrine first propounded by Frederick J. Powell.[12] Under the

---

[5] For example, the corporate veil doctrine can be pierced to allow the plaintiff to assert jurisdiction over the parent corporation which has no business presence in a state except for its subsidiary. In such a case, veil piercing does not immediately result in shareholder liability. *See* Radaszewski v. Telecom Corp., 981 F.2d 305 (8th Cir. 1992). For other veil piercing cases that did not involve imposition of shareholder liability, see Cannon Mfg. Co. v. Cudahy Packing Co., 267 U.S. 333 (1925); *see also* Pauley Petroleum, Inc. v. Cont'l Oil Co., 239 A.2d 629 (1968). This Article will focus on instances of veil piercing which directly implicate limited liability.

[6] Liability may be imposed on others associated with the corporation through various tort doctrines such as vicarious liability. *See* Robert B. Thompson, *Unpacking Limited Liability: Direct and Vicarious Liability of Corporate Participants for Torts of the Enterprise*, 47 VAND. L. REV. 1, 24-39 (1994) (hereinafter Thompson I).

[7] The doctrine may apply to set aside the separate corporate entity without imposing shareholder liability. For example, the identity of a subsidiary may be ignored to allow a court to exert jurisdiction over a corporate parent. *See, e.g.*, Cannon Mfg. Co. v. Cudahy Packing Co., 267 U.S. 333 (1925); Empire Steel Corp. of Tex., Inc. v. Super. Ct. of L.A. County, 366 P.2d 502 (1961); Pauley Petroleum, Inc. v. Cont'l Oil Co., 239 A.2d 629 (1968). Or the separate personality of a subsidiary may be disregarded to prevent the circumvention of a statute. *See, e.g.*, State v. Swift & Co., 187 S.W.2d 127 (Tex. App. 1945). Therefore, separate corporate personality and limited liability are not necessarily synonymous with each other, and the former can be overridden without implicating the latter. For the purpose of this Article, the discussion will focus on application of the corporate veil doctrine which results in the imposition of shareholder liability. These cases arguably present the most interesting theoretical issues.

---

[8] Laborers' Pension Fund v. Lay-Com, Inc., 580 F.3d 602, 610 (7th Cir. 2009) ("Veil piercing is an equitable remedy governed by state law . . . ."); Liberty Prop. Trust v. Republic Props. Corp., 577 F.3d 335, 340 (D.C. Cir. 2009); Williamson v. Recovery Ltd. P'ship, 542 F.3d 43, 53 (2d Cir. 2008) ("Instead of a firm rule, the general principle guiding courts in determining whether to pierce the corporate veil 'has been that liability is imposed when doing so would achieve an equitable result.'" (quoting William Wrigley Jr. Co. v. Waters, 890 F.2d 594, 601 (2d Cir. 1989))); United States v. Farr, 536 F.3d 1174, 1182 n.4 (10th Cir. 2008) ("Courts and commentators have long referred to the equitable principle of lifting or drawing aside the thin guise that is the veil of corporate entity to expose the actual individuals hiding behind it." (internal quotation marks omitted)); HOK Sport, Inc. v. FC Des Moines, L.C., 495 F.3d 927, 935 (8th Cir. 2007) ("Piercing the corporate veil is a common-law equitable remedy whereby an entity's corporate form is disregarded to prevent an injustice."); Bridas S.A.P.I.C. v. Gov't of Turkmenistan, 447 F.3d 411, 420 (5th Cir. 2006) (referring to corporate veil doctrine as equitable remedy); Dearth v. Collins, 441 F.3d 931, 935 (11th Cir. 2006) (corporate veil doctrine governed by equitable principles); IFC Interconsult, AG v. Safeguard Int'l Partners, LLC, 438 F.3d 298, 312 (3d Cir. 2006) ("Moreover, garnishment actions are distinguishable from veil-piercing because they require a pre-existing contractual or statutory basis for garnishment, whereas veil-piercing is an equitable procedure that involves the creation of liability in spite of the contractual and statutory separateness of corporate entities.") (internal citation omitted); Taylor Steel, Inc. v. Keeton, 417 F.3d 598, 606 (6th Cir. 2005) ("[P]iercing the corporate veil is an equitable remedy, available not where a set list of factors are established but where maintaining the corporate form would work injustice upon an innocent party."). *But see* Wm. Passalacqua Builders, Inc. v. Resnick Developers South, Inc., 933 F.2d 131, 136 (2d Cir. 1991) (concluding that veil piercing is both legal and equitable in nature); *see also* David H. Barber, *Piercing the Corporate Veil*, 17 WILLAMETTE L. REV. 371, 374 (1980) ("A 'totality of the circumstances' rule allows the courts to deal with each case on its own facts but fails to provide entrepreneurs with firm guidelines for avoiding personal liability."); Smetherman v. Wilson, 626 So. 2d 71, 73 (La. Ct. App. 1993) (noting that totality of circumstances approach applies to corporate veil cases).

[9] *See, e.g.*, Sonora Diamond Corp. v. Super. Ct., 99 Cal. Rptr. 2d 824, 836 (Cal. Ct. App. 2000) ("In California, two conditions must be met before the alter ego doctrine will be invoked. First, there must be such a unity of interest and ownership between the corporation and its equitable owner that the separate personalities of the corporation

---

and the shareholder do not in reality exist. Second, there must be an inequitable result if the acts in question are treated as those of the corporation alone."); *Laborers' Pension Fund*, 580 F.3d at 610 (Under Illinois law, "[t]he standard test for piercing the corporate veil is two-pronged. The plaintiff must demonstrate both that there is such unity of interest and ownership between the individual or entity and the corporation that the separate personalities of the corporation and the individual [or entity] no longer exist, and that adherence to the fiction of separate corporate existence would sanction a fraud or promote injustice.") (internal quotation marks omitted); Dursharm v. Elegant Custom Homes, Inc. 302 F. App'x 571, 572 (9th Cir. 2008) (To establish a corporate veil claim, "[p]laintiffs must prove . . . that observance of the corporate form would sanction a fraud or promote injustice").

[10] Barber, *supra* note 8, at 376-77.

[11] Robert B. Thompson, *Piercing the Corporate Veil: An Empirical Study*, 76 CORNELL L. REV. 1036, 1063 (1991) (hereinafter Thompson II). For other lists of factors considered by the courts, see Cathy S. Krendl & James R. Krendl, *Piercing the Corporate Veil: Focusing the Inquiry*, 55 DENV. U. L. REV. 1, 16-17 (1978); Barber, *supra* note 8, at 374-75. These substantive factors are to be distinguished from labels such as instrumentality, alter ego, dummy, or agency. These labels merely encapsulate the conclusions drawn by the courts about the formal integrity of the corporation. They tell us nothing about the precise basis upon which the courts determined that the corporation lacked formal integrity.

[12] This doctrine, as applied by the court in *Lowendahl v. Baltimore & Ohio R.R., Co.*, states that the corporate veil should be pierced if the following three elements are present in a case:

> (1) Control, not mere majority or complete stock control, but complete domination, not only of finances, but of policy and business practice in respect to the transaction attacked so that the corporate entity as to this transaction had at the time no separate mind, will or existence of its own; and
>
> (2) Such control must have been used by the defendant to commit fraud or wrong, to perpetrate the violation of a statutory or other positive legal duty, or a dishonest and unjust act in contravention of plaintiff's legal rights; and
>
> (3) The aforesaid control and breach of duty must proximately cause the injury or unjust loss complained of.

instrumentality doctrine, a corporate veil claim requires the plaintiff to show parent domination of the subsidiary corporation, the presence of improper conduct or purpose, and proximate causation between the defendant's conduct and the plaintiff's loss.[13] Powell further enumerated eleven criteria for determining the existence of shareholder domination and seven situations in which impropriety is said to exist. The criteria focus on the degree of operational and financial independence of the subsidiary corporation, overlap in corporate personnel between the parent and the subsidiary, and the formal integrity of the subsidiary.[14] The seven instances of impropriety are: (1) actual fraud; (2) violation of a statute; (3) stripping the subsidiary of its assets; (4) misrepresentation; (5) estoppel; (6) torts; and (7) other cases of wrong or injustice. The instrumentality doctrine further requires the plaintiff to show that the defendant's conduct proximately caused the plaintiff's loss.[15] Only some states have incorporated this requirement.[16]

---

287 N.Y.S. 62, 76 (N.Y. App. Div. 1936), *aff'd*, 6 N.E.2d 56 (N.Y. 1936).

[13] The instrumentality doctrine was originally conceived to apply in the parent-subsidiary setting, but has been extended by the courts to cases involving individual shareholders.

[14] FREDERICK J. POWELL, PARENT AND SUBSIDIARY CORPORATIONS § 6 (1931). The eleven criteria circumstances are: (1) the parent corporation owns all or most of the capital stock of the subsidiary; (2) the parent and subsidiary corporations have common directors or officers; (3) the parent corporation finances the subsidiary; (4) the parent corporation subscribes to all of the capital stock of the subsidiary or otherwise causes its incorporation; (5) the subsidiary has grossly inadequate capital; (6) the parent corporation pays the salaries and other expenses or losses of the subsidiary; (7) the subsidiary has substantially no business except with the parent corporation, or no assets except the ones conveyed to it by the parent corporation; (8) in the papers of the parent corporation or in the statements of the officers, the subsidiary is described as a department or division of the parent corporation, or its business or financial responsibility is referred to as the parent corporation's own; (9) the parent corporation uses the property of the subsidiary as its own; (10) the directors or executives of the subsidiary do not act independently in the interest of the subsidiary, but take their orders from the parent corporation in the latter's interest; and (11) the formal legal requirements of the subsidiary are not observed.

[15] Krendl & Krendl, *supra* note 11, at 12.

[16] Some of the states that have incorporated this proximate causation requirement include Alabama, Florida, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Tennessee, and Wisconsin. Messick v. Moring, 514 So. 2d 892, 894-95 (Ala. 1987); Johnson Enters. of Jacksonville, Inc. v. FPL Group, Inc., 162 F.3d 1290, 1306 (11th Cir. 1998); Gasparini v. Pordomingo, 972 So. 2d 1053, 1055 (Fla. Dist. Ct. App. 2008); Collet v. Am. Nat'l Stores, Inc., 708 S.W.2d 273, 284 (Mo. Ct. App. 1986); Craig v. Lake Asbestos of Que., Ltd., 843 F.2d 145, 148 (3d Cir. 1988); Scott v. AZL

Many of them have not. For the purpose of this Article, the first prong of the corporate veil doctrine will be referred to as the formalities prong and the second, the impropriety prong.[17]

It should be obvious that the corporate veil doctrine has a putative standard-like quality to it, in contrast to the distinctly rule-based character of limited liability. A more in-depth assessment, however, is needed to substantiate this tentative observation. In a seminal article, Professor Duncan Kennedy proposes three dimensions along which rules and standards can be aligned: formal realizability and open-endedness, generality and particularity, and legal commands governing formalities (formalities directives) and legal commands aiming to deter morally wrong conduct (conduct directives).[18] A legal rule is said to be formally realizable and general and a standard open-ended and specific. The last dichotomy requires a bit of explanation. First, it is probably a misnomer to call it a dichotomy. A spectrum would be more apt; between the two ends of the spectrum lie hybrid formalities-conduct directives. Second, rules and standards are not neatly aligned at the two ends of the spectrum. Conduct can be regulated through the use of either rules or standards. As for formalities, although there is nothing inherent in standards that preclude them from being used to govern formalities, the practical reality is that most formalities directives are rule based. The ensuing

---

Res., Inc., 753 P.2d 897, 900 (N.M. 1988); Lowendahl v. Balt. & Ohio R.R. Co., 287 N.Y.S. 62, 77 (N.Y. App. Div. 1936); Freeman v. Complex Computing Co., Inc., 119 F.3d 1044, 1053 (2d Cir. 1997); Glenn v. Wagner, 329 S.E.2d 326, 331 (N.C. 1985); Belvedere Condo. Unit Owners' Ass'n v. R.E. Roark Cos., 617 N.E.2d 1075, 1077 (Ohio 1993); Cont'l Bankers Life Ins. Co. v. Bank of Alamo, 578 S.W.2d 625, 632 (Tenn. 1979); Consumer's Co-op of Walworth County v. Olsen, 419 N.W.2d 211, 217-18 (Wis. 1988). Federal common law also imposes a similar requirement. United States v. Cordova Chem. Co. of Mich., 113 F.3d 572, 585-86 (6th Cir. 1997).

[17] Impropriety is chosen over unfairness because the former conveys more specifically the focus of the inquiry. While the second prong is generally concerned about unfairness resulting from upholding separate corporate personality, it more specifically looks for evidence of improper conduct or purpose on the part of the shareholders that allows them to be enriched at the expense of outside creditors. Krendl & Krendl, *supra* note 11, at 23 (stating existence of improper purposes as ultimate issue in veil piercing analysis).

[18] Duncan Kennedy, *Form and Substance in Private Law Litigation*, 89 HARV. L. REV. 1685, 1687-94 (1976).

analysis will illustrate the extent to which these dimensions describe limited liability and the corporate veil doctrine.

Formal realizability refers to the ability to be applied clearly and predictably to easily distinguishable factual situations. Limited liability clearly fits this description. The factual situation to which it applies is easily distinguishable: when a corporation has been properly formed. Once that condition is met, limited liability provides that shareholders' maximum exposure to corporate liabilities is confined to the value of their equity investments. It has also been said that legal rules possess an "all-or-nothing" character,[19] which aptly describes limited liability. If a corporation was formed in full compliance with the procedures and requirements stipulated in the statute, its owners benefit from the protection of the rule of limited liability. If the incorporation process was incomplete or otherwise faulty, the shareholders would not enjoy limited liability.

Whether the corporate veil doctrine possesses standard-like open-endedness entails a more nuanced explanation. One of the distinct features of an open-ended standard is that it "refers directly to one of the substantive objectives of the legal order . . . [such as] good faith, due care, fairness, unconscionability, unjust enrichment, and reasonableness."[20] H.L.A. Hart calls this quality of a standard desirability. A standard is desirable because it "refer[s] more or less explicitly to some purpose, goal, entitlement, or value, [is] regarded from some point of view as desirable to maintain, or to adhere to, and so not only as providing an explanation or rationale of the rules which exemplify them, but as at least contributing to their justification."[21] In other words, a standard encapsulates and reinforces desirable goals and values. What are the goals and values pursued by the corporate veil doctrine? According to Professor Ballantine, whether the corporate entity should be set aside "comes down to a question of good faith and honesty in the use of corporate privilege for legitimate ends."[22] A federal court asserted that veil piercing is justified "when the notion of legal entity is used to defeat public convenience, justify wrong, protect fraud, or defend crime."[23] These statements suggest that the corporate veil doctrine requires good faith and fair use of the corporate entity. Dean Clark has referred to this duty of good faith and fairness as principles of truth and respect in a corporation's dealing with its creditors.[24] Good faith, fairness, truth, and respect are some of the substantive objectives of the legal order, or desirable goals or values, as according to Hart.[25]

---

[19] H.L.A. HART, THE CONCEPT OF LAW 261-62 (2d. 1994). Hart actually uses the term "principle", while Kennedy offers a number of possibilities for the opposite of a rule—"a standard or principle or policy"—before settling on standard. Kennedy, *supra* note 18, at 1688.

[20] Kennedy, *supra* note 18, at 1688.

[21] Hart, *supra* note 19, at 260.

Another distinct feature of an open-ended standard is that its application "requires the judge both to discover the facts of a particular situation and to assess them in terms of the purposes or social values embodied in the standard."[26] Hart calls this non-conclusiveness.[27] Determination of a veil piercing claim requires the judge to ascertain facts such as the extent of overlap in corporate personnel, inadequate capitalization, non-observance of corporate formalities, and the degree of shareholder domination of the corporation, and evaluate them in light of the doctrine's underlying values of good faith and fairness. It is not uncommon for an identical fact to be interpreted differently by two courts, producing divergent results.[28] Moreover, concepts such as unity of interest and ownership and inequitable results are highly malleable, especially in light of repeated affirmations by the courts that mere complete share ownership on its own does not justify veil piercing.[29] Complete share ownership does not necessarily result in veil piercing, whereas the veil has been pierced in cases in which the controlling shareholder's stake is less than complete.[30] Whether unity of interest between the corporation and its shareholders is found depends on the totality of circumstances and the court's assessment of them. In short, the corporate veil doctrine is non-conclusive. With its reference to desirable values and its non-conclusiveness, the corporate veil doctrine exhibits the open-endedness of a standard.

---

[22] Henry W. Ballantine, *Separate Entity of Parent and Subsidiary Corporations*, 14 CAL. L. REV. 12, 19 (1925).

[23] United States v. Milwaukee Refrigerator Transit Co., 142 F. 247, 255 (E.D. Wis. 1905).

[24] Robert C. Clark, *The Duties of the Corporate Debtor to its Creditors*, 90 HARV. L. REV. 505 (1977). Clark defines the ideal of truth as a requirement that in connection with transfer of property to others, a debtor cannot tell lies that will lead to non-satisfaction of the creditor's claim. *Id.* at 509. The ideal of respect imposes on the debtor "a moral duty in transferring his property to give *primacy* to so-called legal obligations, which are usually the legitimate, conventional claims of standard contract and tort creditors, as opposed to the interests of self, family, friends, shareholders, and shrewder or more powerful bargaining parties." *Id.* at 510-11. For example, removal of assets by a shareholder from the corporation which leaves its creditors' claims unsatisfied would likely result in veil piercing, and violates the ideal of respect.

[25] Kennedy, *supra* note 18, at 1688. One interesting question about the good faith and fair use of the corporate entity is whether it encompasses efforts to incorporate specifically to limit future liability. A number of courts have observed that this use of the corporate entity is fully legitimate. *See* Radaszewski v. Telecom Corp., 981 F.2d 305, 311 (8th Cir. 1992) ("The doctrine of limited liability is intended precisely to protect a parent corporation whose subsidiary goes broke . . . . We think that the doctrine would largely be destroyed if a parent corporation could be held liable simply on the basis of errors in business judgment."); NLRB v. Fullerton Transfer & Storage Ltd., Inc., 910 F.2d 331, 338 (6th Cir. 1990) ("Virtually all corporations are formed for the purpose of limiting liability. Only where an attempt to limit liability is of a type that typically involves a specific attempt to thwart labor law obligations is the application of the relaxed alter ego principles justified."); Craig v. Lake Asbestos of Que., Ltd., 843 F.2d 145, 149 (3d Cir. 1988) ("[E]vasion of tort liability has never, in itself, been sufficient basis to disregard corporate separateness."); Arnold v. Phillips, 117 F.2d 497, 502 (5th Cir. 1941) ("That it [a corporation] was created to shield the owners from liability beyond the capital set up by the charter does not show an unlawful or fraudulent intent, for that is a main purpose of every incorporation.").

---

[26] Kennedy, *supra* note 18, at 1688.

[27] In addition to non-conclusiveness and desirability, Hart also identified breadth as a characteristic of a standard. The breadth of the corporate veil doctrine is demonstrated by the range of circumstances to which it applies. The doctrine applies to cases involving contractual claims, tort claims, statutory claims, criminal claims, etc. It can be invoked for a variety of purposes, including imposition of shareholder liability, assertion of jurisdiction, and prevention of circumvention of a statute.

[28] *See, e.g.*, Automotriz del Golfo de Cal. v. Resnick, 306 P.2d 1 (Cal. 1957); Zubik v. Zubik, 384 F.2d 267 (3d Cir. 1967); Fisser v. Int'l Bank, 282 F.2d 231 (2d Cir. 1960); Empire Steel Corp. v. Super. Ct. of L.A. County, 366 P.2d 502 (Cal. 1961). In *Resnick*, non-observance of corporate formalities was one of the principal grounds for veil piercing, whereas in *Zubik*, separate corporate personality despite serious failure to comply with corporate formalities. In *Fisser*, the plaintiff's awareness of the subsidiary's financial difficulties effectively defeated its corporate veil claims, while such knowledge was tolerated by the court in *Empire Steel* in ordering the veil to be pierced.

[29] Doughty v. CSX Transp., Inc., 905 P.2d 106 (Kan. 1995) (complete stock ownership alone does not turn subsidiary into parent's instrumentality); Berkey v. Third Ave. Ry. Co., 155 N.E. 58 (N.Y. 1926) ("Stock ownership alone would be insufficient to charge the dominant company with liability for the torts of the subsidiary."); Lowendahl v. Balt. & Ohio R.R. Co., 154, 287 N.Y.S. 62, 72-73 (N.Y. App. Div. 1936) ("Control through mere ownership of a majority of even all the capital stock . . . will not in and of itself predicate liability."); Elenkreig v. Siebrecht, 144 N.E. 519, 521 (N.Y. 1924) (affirming separate corporate personality of one-man corporations so long as they are "doing business as permitted by the laws of this State.").

[30] Thompson II, *supra* note 11, at 1055. Thompson found that the courts have pierced the veil in 46% of the cases in which the corporation at issue had two to three

A rule is said to be general and a standard specific. Limited liability is clearly of general application.[31] It applies to all corporations regardless of the number of shareholders, the business in which they operate, or even whether they have business operations at all or merely function as a shareholding parent within a corporate group. In comparison, the corporate veil doctrine has a much narrower scope of application. It only concerns a small subset of corporations exhibiting characteristics that suggest a lack of independence and economic substance. Therefore, limited liability and the corporate veil doctrine fit into the generality-particularity dichotomy of rules and standards posited by Kennedy. To sum up, limited liability is formally realizable and general, while the corporate veil doctrine is open-ended and specific in application. The extent to which limited liability and the corporate veil doctrine govern issues of formality or aim to deter immoral conduct will be dealt with subsequently.

B. The Practical Implications of the Formal Rule-Standard Dichotomy

It has been said that rules create legal certainty and promote investments. By informing parties of the legal consequences of their action *ex ante*, rules allow them to plan their affairs in advance.[32] In contrast, standards discourage commercial activities.[33] The open-ended nature of legal

---

shareholders, and 35% of the cases in which the corporation was a close one but had more than three shareholders.

[31] Kennedy observes that rules need not be of general application. Nor are standards necessarily particular in scope. However, in reality, general rules and particular standards tend to be the norm. Kennedy, *supra* note 18, at 1688.

[32] Kennedy, *supra* note 18, at 1698 ("If an actor knows that the use of a formality guarantees the execution of his intentions, he will do things that he would not do if there were a risk that the intention would be defeated.").

[33] Kennedy further added two reasons that standards retard investment. First, the attendant legal uncertainty "if the judge is at large in finding intent, rather than

standards and the balancing that is involved in their application means that legal consequences can only be ascertained *ex post.* This will make it more difficult for parties to plan their business activities. Within the context of limited liability and the corporate veil doctrine, the implication of these assertions is that limited liability promotes investment while the corporate veil doctrine deters it. Limited liability does exhibit the investment incentive effects of rules. By capping investor exposure to corporate liabilities, it enhances the flow of capital to corporations. There is a substantial body of legal scholarship devoted to explaining how limited liability encourages investment and serves as the foundation of the capital markets.[34] The definiteness of the rule spurs commercial activities, while externalizing the costs of business failure to counterparties that are unable to negotiate additional protection for themselves.

Commentators have argued that at least for contractual creditors, there is little externalization of uncompensated default risks because the creditor can always negotiate for extra credit protection or a higher rate of return.[35] This is in keeping with Kennedy's idea that the formal realizability of legal rules leads actors in society to "invest in formal proficiency", i.e., adjust their behavior in light of the rules.[36] However, the kind of adjustment that is required of the creditors under limited liability is more difficult and complicated than the kind of investment in formal proficiency

---

bound to respond mechanically to ritual acts like sealing, will reduce the payoff that can be expected from being careful." *Id.* at 1698. Second, the incentive to take care is lowered if "the judge may bail you out if you blunder." *Id.*

[34] Paul Halpern, Michael Trebilcock & Stuart Turnbull, *An Economic Analysis of Limited Liability in Corporation Law*, 30 U. TORONTO L.J. 117 (1980); Frank R. Easterbrook & Daniel R. Fischel, *Limited Liability and the Corporation*, 52 U. CHI. L. REV. 89 (1985); Richard A. Posner, *The Rights of Creditors of Affiliated Corporations*, 43 U. CHI. L. REV. 499 (1976); Henry Hansmann & Reinier Kraakman, *The Essential Role of Organizational Law*, 110 YALE L.J. 387 (2000) (hereinafter Hansmann & Kraakman I); Joseph A. Grundfest, *The Limited Future of Unlimited Liability: A Capital Markets Perspective*, 102 YALE L.J. 387 (1992).

[35] Easterbrook & Fischel, *supra* note 34, at 105; Larry E. Ribstein, *The Deregulation of Limited Liability and the Death of Partnership*, 70 WASH. U. L. Q. 417, 442 (1992) (asserting that contractual creditors, including employees, adjust their rates of return to incorporate increased default risks).

[36] Kennedy, *supra* note 18, at 1698.

alluded to by Kennedy. All that a party needs to do to invest in formal proficiency under contract law, as explained by Kennedy, is to familiarize themselves and comply with the rules to make sure that their contracts will be recognized by the courts. The only significant obstacle to this type of formal proficiency is arguably the time and the resources needed to acquire it. The necessary "formal" adjustment under limited liability entails negotiation by the creditors for a higher rate of return or credit protection from the corporation in the form of shareholder personal guarantees or pledge of security. With respect to formal proficiency under contract law, both parties strive for the same outcome and are equally interested in creating a legally binding contract. Meanwhile, the shareholders of the corporation have every incentive to resist the outside creditors' requests. Even if the creditor has the time and the resources to attempt the negotiation, its lack of bargaining power vis-à-vis the corporation may render its effort futile. This is very likely to be the case for employees and to a lesser extent, trade creditors.[37] With respect to tort victims, there is effectively no opportunity for them to negotiate for any formal adjustments, as they almost never know the identity of the tortfeasor in advance. The externalization of accident risks is unmistakable.[38] Therefore, similar to Kennedy's critique of legal rules, limited liability promotes investments through greater *ex ante* certainty but not without imposing substantial costs on those which are unable to acquire formal proficiency.

The central critique of the corporate veil doctrine from the previous paragraph, which remains unaddressed, is whether it discourages commercial activities. The obvious argument is that given the incentive effects of limited liability on corporate investment, the corporate veil doctrine, by overriding limited liability, must inhibit investment. The extent to which this argument is true depends on the circumstances under which the courts have invoked the doctrine. By expressly requiring improper conduct or injustices in order to pierce the veil, it is unlikely that the corporate veil doctrine will deter *bona fide* investors. One may argue that what the courts deem to be improper conduct or injustices are so vague that potential investors are bound to be put off. This argument is unlikely to be valid for a number of reasons. First, while it is admittedly true that the class of conduct which the courts have deemed improper has been broad, few would disagree that the conduct is clearly objectionable. The corporate veil decisions should give clear indications of the scope of permissible conduct that will preserve limited shareholder liability. Second, as Thompson's survey shows, the U.S. courts have never pierced the veil of a public corporation.[39] Setting aside the issue of whether this disparate treatment is theoretically justifiable, the practical reality is that the type of corporation that is most likely to attract passive investments has not been affected by the doctrine.[40] Lastly, and most importantly, liability under the corporate veil doctrine is only imposed on the shareholder who is responsible for the improper conduct. Passive shareholders are almost never implicated, even in close corporations.[41] The corporate veil doctrine will not deter passive investments. As for the culpable shareholders, there should be little disagreement that they should be deterred.

---

[37] Philip I. Blumberg, *Limited Liability and Corporate Groups*, 11 J. CORP. L. 573, 619-20 (1986); Posner, *supra* note 34, at 508, 522-23; Halpern et al., *supra* note 34, at 149-50.

[38] Halpern et al., *supra* note 34, at 145-47; David W. Leebron, *Limited Liability, Tort Victims, and Creditors*, 91 COLUM. L. REV. 1565, 1584 (1991); Henry Hansmann & Reinier Kraakman, *Toward Unlimited Shareholder Liability for Corporate Torts*, 100 YALE L.J. 1879, 1883 (1990) (hereinafter Hansmann & Kraakman II).

The last argument, in particular, points to an important distinction between the corporate veil doctrine and a general rule of unlimited liability. While some commentators have argued that any difference between the two is largely rhetorical, the two in fact operate quite differently.[42] As just mentioned, the corporate veil doctrine only applies to the culpable shareholder, while a general rule of unlimited liability would implicate all shareholders. Moreover, unlimited liability would be a default rule of corporation law that sets out the parties' liability positions *ex ante*, while the corporate veil doctrine intervenes *ex post* to alter these positions. An *ex ante* rule generally has greater impact over parties' conduct than an *ex post* rule, not to mention an *ex post* standard. While one can try to improve the formal realizability of the corporate veil doctrine, it is unlikely to achieve the effectiveness of an *ex ante* rule. The corporate veil doctrine was never intended to be a surrogate for unlimited liability. It is only meant to be a circumscribed exception to the general rule.

---

[39] Thompson II, *supra* note 11, at 1055.

[40] Investors in close corporations are unlikely to be completely passive in the sense of having no relationship with the controlling shareholders whatsoever. These investors are likely to be family members or friends of the controlling shareholders, which should give them somewhat better access to information and a greater ability to monitor the management. They should have fewer concerns about being subject to limited liability because of conduct by wayward management.

[41] Thompson II, *supra* note 11, at 1056.

[42] Hansmann & Kraakman II, *supra* note 38, at 1932.

One further argument that has been made about the deterrent effects of standards on investments is that parties will have reduced incentives to take care if they know that they will be bailed out by judicial intervention. Transposed in the corporate veil context, the argument would be that if creditors know that they can ask the judge to impose shareholder liability *ex post*, they will make less effort to investigate the creditworthiness of the corporation *ex ante*. That clearly would be undesirable. However, there is little evidence to suggest that this should be a serious concern. The courts have repeatedly said that contractual parties should be less entitled to veil piercing in light of their *a priori* opportunity to investigate and negotiate.[43] Moreover, the courts have often refused to bail out contractual parties who failed to obtain adequate *ex ante* credit protection when given the opportunity to do so.[44] In conclusion, while limited liability exhibits the general investment incentive effects of a rule, it externalizes uncompensated default risks and accident risks of a corporation to outside creditors. Meanwhile, if judiciously enforced, the corporate veil doctrine does not retard investments.

---

[43] Southeast Tex. Inns, Inc. v. Prime Hospitality Corp., 462 F.3d 666, 679 (6th Cir. 2006) ("Viewing this case from its proper perspective—a consensual contractual agreement between sophisticated parties—we are not persuaded that Southeast has pleaded extraordinary circumstances that justify piercing May-Ridge's corporate veil."); Williams v. Midwest Employers Cas. Co., 34 F. App'x 152 (5th Cir. 2002) ("[I]n contract cases plaintiff chooses to rely solely on obligation of corporation without any additional guarantees from its shareholders."); Birbara v. Locke, 99 F.3d 1233, 1238 (1st Cir. 1996) ("Several courts and commentators have suggested that it should be more difficult to pierce the veil in a contract case than in a tort case . . . . We have found no Massachusetts Supreme Judicial Court case applying the veil piercing doctrine in a contract case.") (internal citations omitted); Perpetual Real Estate Servs., Inc. v. Michaelson Props., Inc., 974 F.2d 545, 550 (4th Cir. 1992) ("Courts have been extraordinarily reluctant to lift the veil in contract cases, such as this one, where the creditor has willingly transacted business with the corporation.") (internal quotation marks omitted).

C. The Corporate Veil Doctrine as a Hybrid Formalities-Conduct Directive

The third dimension of the rule-standard dichotomy is whether a legal command is used to encourage members of society to conform to certain legal formalities or whether the command is used to deter immoral conduct, or a mixture of both. As mentioned earlier, this dimension consists more of a spectrum than a dichotomy as there are legal commands that implicate both. Moreover, this dimension does not align neatly along rules and standards. At least conduct directives can be formulated as either rules or standards.[45]

It has been said that formalities directives serve cautionary and evidentiary functions by ensuring that "the parties know what they are doing, and . . . that the judge will know what they did."[46] Moreover, "[f]ormalities are premised on the lawmaker's indifference as to which of a number of alternative relationships the parties decide to enter."[47] Lawmakers are presumably indifferent about whether some private parties use a contract to govern their business transaction. The purpose of conduct directives is to reduce the incidence of morally objectionable conduct in society. These directives indicate a clearly expressed preference of the lawmaker against such conduct. These properties of formalities directives and conduct directives will assist in the classification of limited liability and the corporate veil doctrine along the spectrum of directives. An attempt will be made to determine the extent to which they serve evidentiary and cautionary functions and deter immoral conduct. This classification exercise is important because it will improve our understanding of the precise nature and purpose of limited liability and the corporate veil doctrine as legal commands.

---

[44] Bell Oil & Gas Co. v. Allied Chem. Corp., 431 S.W.2d 336 (Tex. 1968) (court refused to pierce veil noting that creditor received the promise it had bargained for); Moore & Moore Drilling Co. v. White, 345 S.W.2d 550 (1961) (court refused to pierce veil since creditor had prior opportunity to investigate corporation's financial conditions); Fisser v. Int'l Bank, 282 F.2d 231 (2d Cir. 1960) (court refused to pierce veil since creditor knew subsidiary was undercapitalized and did not ask for guarantee from parent).

[45] Meanwhile, formalities directives are usually in the form of rules.

[46] Kennedy, *supra* note 18, at 1691.

[47] *Id.*

Limited liability cannot be appropriately classified as a formalities directive. There are a number of reasons for this. First, it serves neither cautionary nor evidentiary functions. It is not a rule governing the procedures of incorporation so that business people know that they have adopted a corporation as the legal form of their business venture, or so that the judge knows that these co-venturers intended to conduct their business in the form of a corporation. Second, it cannot be said that the lawmaker is indifferent between alternative liability rules. The legislature has indicated a clear preference for protecting shareholders at the expense of creditors by choosing limited liability as the default rule. This is because once the default rule has been set, it takes a considerable amount of bargaining power and incentives on the part of the creditor to alter the rule through negotiation. Contrary to what has often been argued, it is not as easy to negotiate to override the default rule as it is to preserve it. It takes considerably more time and resources to negotiate against the default rule of the legal regime. Therefore, the adoption of limited liability indicates a conscious policy choice on the part of the lawmakers to favor shareholders and hence to encourage entrepreneurship and capital flow to corporations. Nor does limited liability aim to deter immoral conduct. It merely prescribes the liability positions of the shareholders and creditors following successful incorporation. The best way to characterize limited liability is as a hybrid formalities-conduct directive. Despite the lack of evidentiary and cautionary functions, limited liability is to some extent concerned with formalities. A corporation that was not properly formed may put the limited liability protection for its owners in jeopardy. Limited liability also has a conduct element, except that rather than deterring immoral conduct, it is more aptly viewed as being concerned with encouraging desirable conduct—that of business activities and investments in corporations.

With respect to the corporate veil doctrine, the intuitive response is that it primarily aims to deter immoral conduct, such as the use of the corporate entity to perpetrate fraud or to evade legal obligations. This is manifested in the second prong of the two-part test adopted by many states and the impropriety prong of the instrumentality doctrine. Deterrence is meant to be achieved by holding shareholders responsible for corporate liabilities should they be found to have exploited the privileges of incorporation in an unfair or dishonest manner.[48] This characterization is, however, incomplete as the corporate veil doctrine is also concerned with corporate formalities. One of the criteria identified by Thompson as being frequently applied by the courts is non-observance of corporate formalities. In fact, in the context of veil piercing, the notion of corporate formalities can be understood as going beyond such procedural matters as whether shares were properly issued or whether shareholders' meetings have been properly convened to a broader range of indicia for the formal integrity of the corporation. In this sense, the criteria enumerated by Thompson—lack of substantive separation between the shareholders and the corporation, inadequate capitalization of the corporation, non-observance of corporate formalities, shareholder domination, and overlapping of corporate personnel and management—can be said to be matters of formalities. They are formalities that connote the extent to which the corporation functions autonomously from its shareholders and deserves to have its separate legal personality respected. They signify the formal integrity of the corporation. The courts will only accord limited liability protection to the putative shareholders if these expanded formalities requirements are met.

By piercing the veil when these requirements are not met, the courts have tacitly extended the formalities of incorporation beyond what is specified in the various states' corporation statutes. This extension in fact can be quite substantial, and is most vividly illustrated by the inadequate capitalization cases. Most states have no minimum capital requirement for corporations.[49] Only a handful of jurisdictions, such as District of Columbia, still maintain a minimum capital requirement of some kind.[50] However, by piercing the veil on the grounds of inadequate capitalization, the courts have raised the *de facto* minimum capital requirement from a non-existent one under the statutes to a substantial one requiring corporations to have "sufficient funds to meet obligations to those who must deal with it."[51] The corporate veil doctrine has also extended the temporal dimension of the incorporation formalities. Most of the statutory incorporation procedures only apply at the time of incorporation. There are few ongoing obligations on the shareholders imposed by the state corporation law statutes. The corporate veil doctrine imposes an ongoing obligation on the shareholders not to commingle their assets with the corporation's, not to dominate the daily operations of the corporation, and not to permit overlap in corporate personnel.

---

[48] The extent to which the corporate veil doctrine achieves effective deterrence will be dealt with subsequently.

These judicially expanded formalities do not completely fit Kennedy's description of formalities directives. They do not serve any evidentiary function. The issue in these corporate veil cases is almost never about conveying the intention of incorporation to the judge. There is usually no mistake about the putative shareholders' intention. These formalities requirements do serve a cautionary function in some instances. To the extent that creditors of a corporation are confused about the identity of their counterparty—i.e., whether it is a corporation or its shareholders in their individual capacity—because of non-compliance with these formalities requirements, the courts have been willing to find misrepresentation by the corporation and impose shareholder liability.[52] The rationale is that non-compliance with these requirements constitutes a failure on the part of the putative shareholders to convey to the counterparty that it is transacting with a corporation. Veil piercing provides incentives to the putative shareholders to be forthcoming to the creditors about the precise relationships between the shareholders and the corporation.[53]

---

[49] For example, the Model Business Corporation Act used to require a minimum capital contribution of $1,000. MODEL BUS. CORP. ACT ANN. § 51 (1966). That was abolished in 1969.

[50] D.C. CODE ANN. § 29-101.47(a)(6).

[51] Consol. Sun Ray, Inc. v. Oppenstein, 335 F.2d 801, 806-07 (8th Cir. 1964) (emphasis removed). *But see* Walkovszky v. Carlton, 223 N.E.2d 6, 10 (N.Y. 1966) (suggesting that minimum liability insurance requirement imposed by legislature represented full extent of protection legislature intended taxicab passengers to enjoy).

Kennedy argues that lawmakers are indifferent about compliance with formalities directives. This suggests that judges should be equally neutral about the non-compliance with these judicially imposed requirements.[54] While this may sound surprising, it is consistent with the weight of the case law that requires a finding of improper conduct or purpose in order for the corporate veil to be pierced.[55] The courts have repeatedly held that evidence indicating unity of interest and ownership between the shareholders and the corporation alone is insufficient to justify veil piercing.[56] A close examination of the case law suggests that the ultimate motivation behind judicial intervention in the corporate veil cases is the existence of fraud or injustice, and not whether corporate formalities, broadly or narrowly defined, have been complied with.

---

[52] Doughty v. CSX Transp., Inc., 905 P.2d 106 (Kan. 1995) (plaintiff worker's veil piercing rejected partly because he had no mistake about identity of his employer); Bell Oil & Gas Co. v. Allied Chem. Corp., 431 S.W.2d 336 (Tex. 1968) (court refused to pierce veil as there was no misrepresentation and mistake about identity of corporation); Zaist v. Olson, 227 A.2d 552 (Conn. 1967) (court pierced veil partly on grounds of confusion of identity between shareholder and corporation); Black & White, Inc. v. Love, 367 S.W.2d 427 (Ark. 1963) (two corporations sharing dispatch and other functions reasonably led plaintiff to assume that they were same corporation); Oriental Investing v. Barclay, 64 S.W. 80 (Tex. App. 1901) (plaintiff workers did not know of existence of subsidiary).

[53] Posner has similarly argued for a misrepresentation exception to the general rule of limited liability. According to him, misrepresentation increases the costs which a creditor must incur to investigate and understand the corporation's finances with no corresponding social benefits. Such expenditure of information costs is socially wasteful and can be obviated by creating a misrepresentation exception to limited liability. Posner, *supra* note 34, at 520-21.

[54] Kennedy, *supra* note 18, at 1691 ("Formalities are premised on the lawmaker's indifference as to which of a number of alternative relationships the parties decide to enter."). Kennedy's discussion assumes that the legislature crafts the formalities rules, whereas in the context of the corporate veil doctrine, this is undertaken by the judiciary.

[55] In this Article, improper purpose and conduct is meant to encompass all the usual circumstances under which the courts pierce the veil, such as to prevent fraud, illegality, injustice, a contravention of public policy, or circumvention of a statutory prohibition.

[56] *See supra* text accompanying note 9.

Given its dual concerns with conduct and formalities, the corporate veil doctrine is more appropriately classified as falling within "an intermediate category of legal institutions that partakes simultaneously of the nature of formalities and of rules designed to deter wrongdoing."[57] Absent an improper purpose or conduct, the non-observance of corporate formalities will not result in shareholder liability. However, by making such non-observance a prerequisite for veil piercing, the courts have clearly expressed a desire to encourage shareholders to respect the formal integrity of their corporations. The foregoing analysis thus confirms the corporate veil doctrine as a specific, open-ended standard that encourages compliance with corporate formalities and aims to deter immoral conduct.

What is most didactic about the classification of the corporate veil doctrine as a hybrid formalities-conduct directive is Kennedy's inclusion in the same category legal directives which he claims to evince an ambiguity of purpose. Examples of such directives are rules defining liabilities and assessing damages in contract and tort law. Kennedy argues that these rules may be viewed as harboring moral ambivalence about a contractual breach or a tort so long as the damages are paid.[58] Equally plausibly, they may be characterized as taking a determinedly moralistic view and desiring to deter such conduct.[59] This claim of ambiguity of purpose leads one to ponder the purpose of the corporate veil doctrine. The important question, which does not seem to have been raised, let alone answered, by previous commentators is: what is the objective of the corporate veil doctrine? Does it aim to punish, deter, or compensate? Is it indifferent toward improper conduct and disregard of corporate formalities so long as shareholders bear full responsibility for the corporate liabilities? These questions will be answered in the next two sections.

---

[57] Kennedy, *supra* note 18, at 1692.

[58] *Id.* at 1694.

[59] *Id.*

II. THE SUBSTANTIVE INDIVIDUALISM-ALTRUISM DICHOTOMY OF THE CORPORATE VEIL DOCTRINE

A. Limited Liability as an Individualistic Rule and the Corporate Veil Doctrine as an Altruistic Standard

The relevance of Kennedy's framework to understanding the relationship between limited liability and the corporate veil doctrine is not confined to the formal dimension. His assertion of the correspondence between the rule-standard classification and the individualism-altruism dichotomy sheds light on the objective of the corporate veil doctrine. One of the greatest insights in his analysis of the choice between rule and standard is the fact that this choice reflects a more fundamental conflict between two opposing philosophies or ways of life: individualism and altruism.[60] He explains individualism as

> [T]he making of a sharp distinction between one's interests and those of others, combined with the belief that a preference in conduct for one's own interest is legitimate . . . . The form of conduct associated with individualism is self-reliance. This means an insistence on defining and achieving objectives without help from others . . . .[61]

Altruism is in turn defined as "the belief that one ought *not* to indulge a sharp preference for one's own interest over those of others. Altruism enjoins us to make sacrifices, to share, and to be merciful."[62] Kennedy observes that there is a close parallel in the arguments made by a lawyer for the strict adherence to legal rules and the arguments in defense of individualism.[63] Likewise, there is a distinctly altruistic undertone in the advocacy for standards.[64] The ensuing analysis demonstrates the extent to which limited liability and the corporate veil doctrine reflect individualistic and altruistic beliefs as articulated by Kennedy.

---

[60] Kennedy, *supra* note 18, at 1745-47.

[61] *Id.* at 1713.

[62] *Id.* at 1717.

[63] *Id.* at 1738.

[64] *Id.* at 1740. More relevant to this Article is the association between arguments for rules and laissez-faire economics, and advocacy for standards and liberal economic views, which are more receptive of the imposition of altruistic duties on the well-to-do in society. *Id.* at 1740-51.

There are three specific ways in which rules encapsulate individualism. First, rules are meant to be non-interventionist or at least non-resulted oriented.[65] This is consistent with the requirements of an individualistic society, in which the state takes a back seat and allows individuals the maximum room to take private initiatives and to realize their potential. Second, rules, especially those governing the formalities of transactions, encourage members of society to acquire formal proficiency by punishing non-compliance with non-enforcement of transactions. Similarly, individualism motivates members of society to acquire substantive proficiency, i.e., being economically productive and self-sufficient, by vowing not to bail them out if they fail or lay idle. Third, by increasing legal certainty, rules encourage transactions and exchanges, which are dearly valued by individualists and are presumptively good for society.[66]

It should be obvious that limited liability is individualistic in nature. Limited liability itself emerged in the nineteenth century as an economic policy tool to promote industrialization.[67] As the capital need of businesses grew, state legislatures were under pressure to introduce limited liability to help corporations attract investments from passive shareholders.[68] Limited liability was first adopted in quick succession by the northeastern states, which faced regulatory competition with one another and did not want to lose businesses to other states.[69] In England, limited liability was not adopted until the Limited Liability Act of 1855, and not without a bitter struggle which saw proponents of laissez-faire economics emerge triumphant.[70] In fact, alongside negligence, limited liability is one of those legal rules that encapsulated nineteenth-century laissez-faire economics.[71] Limited liability is probably the quintessential non-interventionist legal rule. Its nature is facilitative instead of regulatory by setting out the default liability positions of the shareholders and the creditors. The main purpose of limited liability, as identified by many scholars, is to encourage equity investments in corporations and facilitate transactions between corporations and outside creditors.[72] It has been argued that limited liability is crucial to the efficient functioning of the capital markets and to corporations' ability to attract passive investments.[73] Limited liability also reduces the need for monitoring by contractual creditors by permitting the compartmentalization of corporate and shareholder assets.[74] Lastly, there is a distinct undertone of self-reliance in limited liability. The rule enjoins creditors of a corporation to use their wherewithal to fend for themselves. It encourages them to spend time to investigate the corporation's financial position and to bargain for additional credit protection. The incentive to undertake precautions *ex ante* is provided by the refusal of the law to bail out *ex post* creditors which have neglected to do so, were mistaken in their effort, or dissatisfied with their bargain.[75] All these characteristics point to a distinctly individualistic nature of limited liability.

---

[65] *Id.* at 1741. Kennedy disagrees with the assertion that rules are non-interventionist and argues that both rules and standards are interventionist, simply to different degrees. *Id.* at 1747-48.

[66] *Id.* at 1744. Moreover, Kennedy points out rules have been traditionally associated with classical laissez-faire economics of the nineteenth century. *Id.*

[67] E. Merrick Dodd, *The Evolution of Limited Liability in American Industry: Massachusetts*, 61 HARV. L. REV. 1351, 1352 (1948).

[68] *Id.* at 1366-73; Jonathan M. Landers, *A Unified Approach to Parent, Subsidiary, and Affiliate Questions in Bankruptcy*, 42 U. CHI. L. REV. 589, 617 (1975).

[69] Blumberg, *supra* note 37, at 590-94.

The corporate veil doctrine is altruistic in the sense that it softens the harshness of the highly self-reliant world envisioned by limited liability. In this world, a contractual creditor has the opportunity to bargain with the corporation *ex ante* and should not be allowed to overturn the bargain *ex post*. Even with respect to fraud and misrepresentation, the individualist may argue that the creditor should and could have anticipated such occurrences and bargained for protection accordingly. Fraud and misrepresentation is common after all, and creditors should be aware of them in advance. Despite its lack of opportunity to engage in prior bargain, a tort victim is no more deserving of special favor from the law as it can protect itself with insurance coverage. The corporate veil doctrine takes exception from this world-view and offers assistance to those who have been unfairly prejudiced by shareholder misconduct. It rejects the individualist stance that every corporate counterparty is in a position to protect itself from shareholder abuses. Nor will the law tolerate sharp practices by unscrupulous shareholders the full brunt of which is most likely to be borne by the weak. In this sense, the corporate veil doctrine takes on a somewhat paternalistic tone, which Kennedy identifies as a characteristic of altruism.[76] The doctrine undertakes a highly particularized inquiry into the specific facts of each case to determine when assistance should be extended. This requires standard-based legal intervention, which, of course, is an altruistic tool.[77] Moreover, while the individualist largely sees the role of the law as facilitative, the altruist is more ready to use it to achieve regulatory objectives. In the context of the corporate veil doctrine, the "regulatory" objective pursued is the prevention of abuses of the corporate entity, especially at the expense of other members of society.

---

[70] Halpern et al., *supra* note 34, at 118.

[71] MORTON J. HORWITZ, THE TRANSFORMATION OF AMERICAN LAW: 1780 to 1860, 89-94, 99-101 (1977).

[72] Halpern et al., *supra* note 34, at 135-38; Fischel & Easterbrook, *supra* note 34, at 95-97; Thompson I, *supra* note 6, at 17-18; Posner, *supra* note 34, at 507-09.

[73] Halpern et al., *supra* note 34, at 135-38; Fischel & Easterbrook, *supra* note 34, at 95-97.

[74] Hansmann & Kraakman I, *supra* note 34, at 424.

[75] William O. Douglas & Carrol M. Shanks, *Insulation from Liability through Subsidiary Corporations*, 39 YALE L.J. 193, 211 (1929) ("The very reasonable question must be met and answered why one who contracted with the subsidiary and received the promise which he bargained for but who has been disappointed in the fulfillment by the subsidiary of its commitment should then be allowed to look to the parent. As a matter of contract it is evident he may not.").

In short, not only do limited liability and the corporate veil doctrine demonstrate the respective formal characteristics of rules and standards, they also conform to the individualist-altruist paradigm posited by Kennedy. Both in formal and substantive senses, limited liability and the corporate veil doctrine fit into Kennedy's formal and substantive dichotomies of private law litigation.

---

[76] Kennedy, *supra* note 18, at 1736.

[77] *Id.* at 1740.

B. Objectives of the Corporate Veil Doctrine

A determination that the corporate veil doctrine is altruistic in nature does not definitively address the question of what its main objective is, which will be answered in the ensuing inquiry. Two observations are in order before embarking on this inquiry. First, the purpose of the inquiry is not to ascribe a single objective to the doctrine, but to ascertain its principal, or at least its most important one. Given the variety of corporate veil cases, it is probably impossible to arrive at a single objective that encapsulates all the cases. In fact, not only can a multitude of objectives be found among the cases, they can be sometimes detected in one single case. These objectives are not mutually exclusive; the courts have expressed a desire both to compensate and to deter or punish within one case. Second, there are few explicit references to the objectives of the doctrine in the cases. What the courts have repeatedly said is that the corporate veil doctrine is meant to prevent injustices.[78] However, a review of the case law shows that the courts have cited prevention of injustices not so much as an overarching justification for the doctrine as a necessary condition for its invocation.[79]

A determination of the principal objective of the doctrine entails an understanding of the interaction between the doctrine itself and the underlying causes of action. Or to frame the inquiry differently, one needs to determine whether the doctrine serves any independent purpose, or merely reinforces the objectives of the underlying contractual or tort claims—which may be compensatory or deterrent, or even punitive if the corporation is subject to punitive or liquidated damages. The answer to this question depends on whether the doctrine expands the underlying contractual or tortious liability beyond its original scope or merely preserves it.

---

[78] HOK Sport, Inc. v. FC Des Moines, L.C., 495 F.3d 927, 936 (8th Cir. 2007); *In re* Blastein, 192 F.3d 88, 100 (3d Cir. 1999); *In re* Schimmelpenninck, 183 F.3d 347, 357 n.21 (5th Cir. 1999); Green v. Windsor Mach. Prods., Inc., 173 F.3d 591, 594 (6th Cir. 1999); United States v. Cordova Chem. Co. of Mich., 113 F.3d 572, 585 (6th Cir. 1997); McCall Stock Farms, Inc. v. United States, 14 F.3d 1562, 1567 n.4 (Fed. Cir. 1993); NLRB v. Greater Kan. City Roofing, 2 F.3d 1047, 1051 (10th Cir. 1993); Manville Sales Corp. v. Paramount Sys., Inc., 917 F.2d 544, 552 (Fed. Cir. 1990).

[79] For instance, the *Manville Sales* court asserted that "a court may exert its equitable powers and disregard the corporate entity if it decides that piercing the veil will prevent fraud, illegality, injustice, a contravention of public policy, or prevent the corporation from shielding someone from criminal liability." *Manville Sales,* 917 F.2d at 552.

Despite statements by some courts to the contrary[80], the corporate veil doctrine does not expand the liability imposed on the corporation. The amount of liability imposed on the corporation remains the same after operation of the doctrine. It merely ensures that liability is met to the fullest extent possible.[81] This is corroborated by the judicial pronouncement that the doctrine is "an equitable remedy, not a cause of action unto itself, which is used as a means of imposing liability."[82] This statement implies that the doctrine is not an independent basis of liability. It merely transposes the liability, to the extent that it is unmet after corporate assets have been exhausted, from the corporation to the defendant shareholder.

This observation, however, is only partially correct. The primary consequence of veil piercing is of course the imposition of shareholder liability. There is in fact a secondary and often overlooked consequence, which is the concentration of liability. While it is true that the doctrine does not alter the size of the underlying liability, it concentrates the entire liability on one or a few culpable shareholders. So long as the corporation has only one shareholder, this concentration makes no difference. For corporations with more than one shareholder, the extent to which the corporate veil doctrine concentrates liability crucially depends on whether shareholder liability is imposed on and prorated among all the shareholders. This is because for a solvent corporation that is not subject to a veil piercing claim, the impact of liability will be borne by the shareholders proportionate to the value of their shareholdings. Take the example of a corporation with $10 million assets and five equal shareholders. If the corporation is subject to $5 million tort liability, each shareholder's claim to the corporate assets will be reduced proportionately by half, absent alternative arrangements. If the tort liability is increased to $15 million, this proportionate allocation would only be maintained if the excess liability of $5 million were distributed equally among the five shareholders. In other words, the unmet liability must be imposed on the entire body of shareholders and be prorated among them.

---

[80] IFC Interconsult, AG v. Safeguard Int'l Partners, LLC, 438 F.3d 298, 312 (3d Cir. 2006) ("[V]eil-piercing is an equitable procedure that involves the creation of liability in spite of the contractual and statutory separateness of corporate entities." (citation omitted)).

[81] The extent to which liability will be paid in full also depends on the amount of shareholder assets. In a way, every liability is limited in the sense that it is subject to the possibility of judgment-proof defendants, corporate or individual. Leebron, *supra* note 38, at 1575.

[82] Hodak v. Madison Capital Mgmt., L.L.C., Nos. 08-6142, 08-6543, 2009 WL 2903587, at *9 (6th Cir. Sept. 10, 2009); *see also* Peacock v. Thomas, 516 U.S. 349, 354 (1996) ("Piercing the corporate veil is not itself an independent ERISA cause of action, 'but rather is a means of imposing liability on an underlying cause of action.'"); Int'l Fin. Serv. Corp. v. Chromas Tech. Canada, Inc., 356 F.3d 731, 736 (7th Cir. 2004) ("Piercing the corporate veil, after all, is not itself an action; it is merely a procedural means of allowing liability on a substantive claim, here breach of contract.").

A review of the corporate veil cases, however, suggests that this is not the way in which the corporate veil doctrine operates.[83] Veil piercing claims are almost invariably targeted at one or a few shareholders responsible for the wrongdoing. They are almost never filed against all the shareholders at the same time. In this sense, the corporate veil doctrine is different from a regime of general unlimited liability that has been proposed by some commentators.[84] A veil piercing court does not prorate liability among the shareholders. Instead, the entire unmet portion of corporate liability is concentrated on the culpable shareholders. The shareholders' incidence of liability is increased through the operation of the corporate veil doctrine. This suggests that the doctrine does not merely reinforce the objectives of the underlying contract or tort law. By altering the incidence of liability, the corporate veil doctrine pursues independent objectives.

---

[83] This author is not aware of any corporate veil cases in which liability was imposed on innocent and culpable shareholders alike.

[84] Hansmann & Kraakman II, *supra* note 38, at 1880-81. Leebron's proposal of unlimited liability, at least with respect to corporate groups, is only limited to majority shareholders of a subsidiary. Leebron, *supra* note 38, at 1612-26. Leebron's proposal is hence more similar to a corporate veil claim.

1. Punishment for Improper Conduct

What remains to be determined is the precise nature of these objectives. There are a number of possibilities, including punishment, deterrence, compensation, and avoiding unjust enrichment. There is evidence that veil piercing has produced punitive effects in some cases. In some of the inadequate capitalization cases, the defendant was subject to a liability that significantly exceeded the shortfall in capitalization identified by the court.[85] The defendant shareholder can be said to have been punished for his failure to capitalize the corporation adequately. There have been other corporate veil cases in which the courts expressed strong disapproval of the defendant's conduct and imposed personal liability as a punishment. For example, in *Bair v. Purcell*, after describing the defendant's conduct as "outrageous and reprehensible," the court pierced the veil to impose hefty personal liability.[86] And of course when pled in a criminal case, veil piercing may further the punitive objectives of the underlying criminal law. However, this is very rare. Of the 1,600 or so cases surveyed by Thompson, only fifteen of them were criminal cases.[87] Therefore, the focus of the discussion will be on the pursuit of punitive objectives in the civil veil piercing cases.

To further a punitive objective in civil cases, a veil piercing court must impose civil damages additional to and far exceeding what is necessary for compensation. The case law shows that this is rarely done by the courts. This can be illustrated with a comparison between the concentration of liability under veil piercing and punitive damages. In civil cases, the courts punish the defendant by imposing punitive damages, which are often multiples of the compensatory award.[88] In corporate veil cases, the same effect can only be achieved if the liability imposed on a shareholder is a multiple of its proportionate share. This concentration effect increases as a particular shareholder's ownership stake decreases in size. For a shareholder with 100% ownership, there will be no concentration effect. For a shareholder with a 20% ownership stake, the liability multiple will be five. A review of the case law shows that veil piercing is most often only applied to shareholders with a controlling stake, which allow the courts to hold those shareholders responsible for the corporation's improper conduct. It is rarely applied to shareholders whose stake is small enough so that the liability multiple will approach that under punitive damages. In other words, veil piercing alone almost never concentrates liability sufficiently to achieve punitive effects. Absent the concurrent imposition of punitive damages, veil piercing is unlikely to achieve significant punitive effects.

---

[85] Gallagher v. Reconco Builders, Inc., 415 N.E.2d 560 (Ill. App. Ct. 1980) (court imposed liability of over $13,000 on defendant for depletion of corporation's $1,000 stated capital); Serv. Iron Foundry, Inc. v. M.A. Bell Co., 588 P.2d 463 (Kan. Ct. App. 1978) (imposing shareholder liability of over $52,000 after noting that balance of corporation dropped from $1,101.96 to negative).

[86] Bair v. Purcell, No. 1:04-CV-1357, 2009 WL 772828, at *15 (M.D. Pa. Mar. 17, 2009). The court's punitive intent was corroborated by its imposition of punitive damages on the defendants. *Id.* at *17 ("Francis Purcell and Norma Purcell intentionally engaged in outrageous, reprehensible conduct and that punitive damages are necessary to punish and deter them and others from engaging in this behavior in the future.").

[87] Thompson II, *supra* note 11, at 1058.

A punitive objective is unsuitable for the corporate veil doctrine for other reasons. First, the class of conduct that justifies veil piercing encompasses a variety of commercial practices. While some cases implicate conduct such as fraud that may justify punitive action, the conduct featured in other cases is often less reprehensible.[89] Punishment will not be an appropriate response in these cases. The ascription of a punitive objective is also inconsistent with the altruistic underpinning of the doctrine, which emphasizes sharing and sacrifice between members of society. Punishment that aims to deter undesirable conduct is a distinctly individualistic response. For these reasons, a punitive rationale is an inappropriate objective of the corporate veil doctrine.

---

[88] For instance, in *Pacific Mutual Life Insurance Co. v. Haslip*, 499 U.S. 1 (1991), the Supreme Court upheld the constitutionality of a punitive damage award which was more than four times the size of the compensatory damages, despite noting that it might be "close to the line." *Id.* at 23. In *BMW of North America, Inc. v. Gore*, 517 U.S. 559 (1996), the Supreme Court struck down as unconstitutionally excessive a punitive damage award that was 500 times the compensatory damages awarded to the plaintiff.

[89] HOK Sport, Inc. v. FC Des Moines, L.C., 495 F.3d 927, 936 (8th Cir. 2007) ("[F]raud is a sufficient, but not necessary, condition for piercing the corporate veil.").

2. Compensation for the Corporate Creditors

The corporate veil doctrine performs a compensatory function. In particular, it helps to restore the full compensatory effects of the underlying contractual or tort damages that otherwise would have been curtailed by limited liability. For example, a contractual creditor whose breach of contract claim is worth $1 million against a corporation with $500,000 assets would have been denied half of its compensation if limited liability were strictly adhered to. The same applies for tort claims. Operation of the limited liability rule would have confined maximum recovery to the sum of the corporation's assets. To the extent that the main function of contractual and tort liabilities is compensatory,[90] the corporate veil doctrine preserves this function.

This picture is incomplete, however, as it ignores the possibility that compensation from a corporation is intended to be subject to an implicit cap under limited liability. Take contractual liabilities as an example. The argument is the well-rehearsed one that when a contractual creditor enters into a transaction with the corporation, it implicitly accepts the fact that its maximum recovery is limited to the corporate assets. If such a creditor desires extra credit protection, perhaps in the form of shareholder personal guarantees, it should negotiate for it. Failure to undertake such negotiation is an implied acceptance of the cap on its maximum recovery imposed by limited liability. This bargain argument has little application for tort victims, who usually do not have the opportunity to negotiate with the tortfeasor *ex ante*. Even then, one may argue that by adopting limited liability, the legislature has made a policy determination that tort damages be subject to a similar cap as well. Potential tort victims are expected to take out insurance to make up for the shortfall. To the extent that these assertions are true, the corporate veil doctrine goes beyond mere restoration of the compensatory function of civil damages by upsetting the tacit bargain that the creditor has made with the corporation and overturning the policy choice made by the legislature.

---

[90] Tort law, of course, is widely understood to perform a deterrent function as well. *See generally* STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW 177-256 (2004).

At least with respect to contractual creditors, it is possible to reconcile veil piercing with the bargain theory proposed above. The reconciliation is premised on the argument that when a contractual creditor transacts with a corporation, one of the implied terms in the contract is that the integrity and autonomy of the corporation will be respected and that the corporation will not be used for improper purposes.[91] In this sense, veil piercing can be viewed as judicial enforcement of the implied terms of the contract. For example, veil piercing based on asset stripping or shareholder misappropriation of corporate assets can be understood as enforcing the implied contractual term that the shareholders will not engage in *ex post* opportunistic conduct.[92] The obvious question is why the parties rely on the courts to fill in the implied terms for them, as opposed to writing them explicitly into the contract themselves. One possible answer is the difficulty with delineating the varieties of *ex post* shareholder opportunism and designing the appropriate protection against them. Instead, the parties choose to rely on the courts to supplement their contracts. This helps to explain why the courts often refuse to pierce the veil when there is evidence suggesting that the contractual creditor foresaw the problem but failed to take any action to protect itself.[93]

---

[91] United States v. Golden Acres, Inc., 702 F. Supp. 1097, 1107 (D. Del. 1988) ("When it agreed to insure the loan to Golden Acres, HUD naturally assumed that Golden Acres would be managed like a normal corporation, with sufficient regard for solvency, corporate formalities, and corporate obligations. Refusal to pierce the corporate veil in this case would be unfair in that it would punish HUD for its misplaced trust and reward defendants for their abuse of the corporate form."); Taylor Steel, Inc. v. Keeton, 417 F.3d 598, 610 (6th Cir. 2005) (noting that even though breaches of contract happen regularly, contractual parties are entitled to expect to have recourse to corporate assets and not to have those assets siphoned off to avoid liability); Laya v. Erin Homes, Inc., 352 S.E.2d 93, 100 (W. Va. 1986) ("[I]n entering into such [contractual] relationships with corporate entities the parties are [generally] entitled to rely upon certain assumptions, one being that the corporation is more than a mere shell—that it has substance as well as form.").

[92] This can also be viewed as an embodiment of the principles of truth and respect proposed by Dean Clark for the dealing between a corporation and its creditors. Clark, *supra* note 24, at 511.

This implied contract explanation for veil piercing, however, suffers from a number of limitations. First, while some courts have refused to pierce the veil in light of evidence that the creditor failed to take precautions against known sources of risk, just as many courts have imposed shareholder liability despite such evidence.[94] If the creditor has expressly considered a potential default risk and has resolved to do nothing about it, judicial intervention cannot be understood as enforcement of implied contractual terms. Instead, what the court does is override the parties' contractual understanding. Second, not every ground for veil piercing can be plausibly characterized as an implied contractual term which the parties would have incorporated in their contract but for the prohibitive transaction costs. For example, a contractual creditor is unlikely to be concerned about the overlap in personnel between a parent corporation and its subsidiary so long as the subsidiary remains profitable and its liabilities are backed up by ample assets. Nor is this creditor likely to negotiate for contractual terms governing the observance of corporate formalities, such as whether shareholder and board meetings are properly convened. This explains why veil piercing criteria such as non-observance of corporate formalities and shareholder domination are usually justified not so much on creditor protection grounds as on an estoppel theory. The theory is that if the shareholder himself does not treat the corporation as one, he cannot be heard to complain when the court sets aside the corporate entity.[95]

Third, this implied contract explanation obviously does not apply to tort claims. There is almost never any *ex ante* negotiation of risk allocation between the corporation and the tort victim. The lack of such negotiation means that veil piercing cannot be explained as judicial fulfillment of the parties' implicit wishes. The conclusion is probably unavoidable that by piercing the veil in tort cases, the courts have overridden the legislature's policy choice. In fact, it is not entirely clear that this explanation applies to contractual creditors of all kinds.[96] One of the basic premises of the explanation is that a contractual creditor can bargain for extra credit protection beyond the background rule of limited liability if it so desires. By piercing the veil, the courts are merely filling in the gaps in the contract and inserting terms that the parties would have wanted to include. However, if the creditor at issue does not have the bargaining power or the incentive to negotiate for additional protection, the function performed by veil piercing is more than gap filling. For instance, an employee is unlikely to have the bargaining power to ask for a personal guarantee for his wages by the shareholders of his employer, let alone to demand contractual protection against *ex post* shareholder opportunism or other kinds of misconduct.[97] It therefore makes little sense to characterize veil piercing on behalf of unpaid employees as enforcement of implied contractual terms. The more plausible explanation is judicial recognition of the employee's inability to protect himself and intervention on his behalf to override the harsh effects of limited liability. The implied contract explanation is arguably only apt for financial creditors and some trade creditors.

---

[93] Bell Oil & Gas Co. v. Allied Chem. Corp., 431 S.W.2d 336 (Tex. 1968); Moore & Moore Drilling Co. v. White, 345 S.W.2d 550 (Tex. Civ. App. 1961); Fisser v. Int'l Bank, 282 F.2d 231 (2d Cir. 1960); *In re* John Koke Co., 38 F.2d 232 (9th Cir. 1930); First Nat'l Bank of Seattle v. Walton, 262 P. 984 (Wash. 1928).

[94] Ampex Corp. v. Office Elecs., Inc., 320 N.E.2d 486 (Ill. App. Ct. 1974); Empire Steel Corp. of Texas v. Super. Ct. of L.A. County, 366 P.2d 502 (Cal. 1961); Automotriz del Golfo de Cal. v. Resnick, 306 P.2d 1 (Cal. 1957); First Nat'l Bank in Canyon v. Gamble, 132 S.W.2d 100 (Tex. Comm'n App. 1939).

[95] Barber, *supra* note 11, at 400; Landers, *supra* note 68, at 621-22.

To sum up, the extent to which the corporate veil doctrine performs a compensatory function depends on the nature of the underlying claim and the identity of the creditor. For a tort victim, to the extent that the legislature has made a conscious policy choice under limited liability to cap tort damages against corporations, the doctrine's effect goes beyond mere compensation.[98] For a contractual creditor, the compensatory effects vary with the type of contractual creditor at issue. For sophisticated financial creditors, one can postulate veil piercing as a compensatory tool by characterizing the judicially imposed requirements as what the parties themselves would have contracted for. These creditors realistically could have bargained for many of the protections mandated by the courts. For many trade creditors and employees, this explanation loses its persuasiveness as most of them would have been unable to obtain those protections through contractual negotiations. Veil piercing performs more than a mere compensatory function for them. In short, while the corporate veil doctrine no doubt performs a compensatory function, compensation is not its principal objective.

---

[96] There are generally three kinds of contractual creditors: financial creditors, trade creditors, and employees. Easterbrook & Fischel, *supra* note 34, at 104-05. Professors Easterbrook and Fischel include consumers as contractual creditors of a corporation as well. There are disagreements regarding the correct classification of consumers as voluntary or involuntary creditors. Blumberg, *supra* note 37, at 618. For the purpose of this article, consumers will be excluded from contractual creditors.

[97] Blumberg, *supra* note 37, at 619-20; Halpern et al., *supra* note 34, at 149-50. Dean Clark has similarly noted that as compared to bank creditors, trade creditors are likely to have less bargaining power and wherewithal to conduct a thorough credit check and negotiate for adequate protection. Clark, *supra* note 24, at 543-45.

3. Deterrence of Future Improper Conduct

Deterrence is another possible objective of veil piercing. It is a particularly fitting one in light of the requirement of improper conduct in the various formulations of the doctrine. In Powell's formulation of the instrumentality doctrine, the types of conduct that qualify as improper conduct include fraud, circumvention of a statutory prohibition, asset stripping, misrepresentation, estoppel, torts, and other cases of wrong or injustice.[99] This is conduct that the law would like to deter.[100] There have been cases in which the courts combined veil piercing with punitive damages with the express purpose to achieve deterrence.[101] In In re *Amberjack Interests*, the court imposed personal liability and punitive damages on the defendants to "deter future conduct of this deplorable nature," which consisted of misappropriation of corporate funds.[102] Similarly, in In re *Fox*, the court combined both personal liability and punitive damages "to deter the defendant from engaging in similar conduct in the future."[103] However, the need to combine shareholder liability with punitive damages to achieve deterrent effects seems to suggest that veil piercing alone does not suffice for deterrence purposes. Moreover, attributing deterrence as the principal objective of the doctrine is complicated by the general judicial consensus that the doctrine is an equitable one whose goal is to prevent injustices.[104] The courts seldom describe the corporate veil doctrine in deterrence terms. Yet the importance of this dearth of references should not be overstated. The courts seldom explicitly acknowledge deterrence as a goal of tort law, but that has not stopped commentators from analyzing and theorizing about the deterrent effects of tort liabilities. What matters is not whether the courts themselves perceive their veil piercing decisions in deterrence terms, but whether these decisions do in fact achieve those effects.

---

[98] The corporate veil doctrine will also restore the deterrent effects of the tort judgment.

[99] POWELL, *supra* note 14, at §§ 6, 13.

[100] With respect to tort liabilities, the corporate veil doctrine helps to preserve the inherent deterrent functions of tort damages. Of course, preservation of the functions of the underlying contractual or tort damages is not the main focus of our inquiry; the focus is whether the doctrine pursues any independent objectives of its own.

[101] In a previous section, there was discussion of courts using punitive damages to punish. Punitive damages of course can both punish and deter. Some courts have

Examination of the possible deterrent effects of the corporate veil doctrine requires a more nuanced understanding of how liability creates deterrence and how this inducement works in the context of the corporate veil doctrine. For the law to have a deterrent effect on undesirable conduct, the penalty must be greater than the benefits the defendant obtains from such conduct. Otherwise, the defendant would be always

---

asserted that the main goal of punitive damages is deterrence. Fischer v. Johns-Manville Corp., 512 A.2d 466, 476 (N.J. 1986).

[102] *In re* Amberjack Interests, Inc., 326 B.R. 379, 394 (Bankr. S.D. Tex. 2005).

[103] *In re* Fox, 232 B.R. 229, 231 (Bankr. D. Kan. 1999). The court noted that the defendants debtor,

> [I]ntentionally concealed his misconduct over several years. He filed fraudulent registrations with the Patent and Trademark Office, deliberately sold inferior aluminum phosphide as Quick-Phos, provided erroneous deposition testimony, falsified multiple invoices, and submitted fraudulent information to both plaintiffs and the court. Further, his trial testimony was hardly the hallmark of truthfulness.

Id.

[104] *See supra* text accompanying note 8.

willing to take a chance with the hope that the conduct will not be detected. If the penalty is the same as the benefit from the undesirable conduct, the defendant will have nothing to lose if it gets caught. In fact, theoretically, the penalty or damages should be set as the benefit multiplied by the reciprocal of the probability of apprehension. For example, if the shareholder's dissipation of corporate assets has a 20% chance of detection, and the dissipated assets are worth $1 million, the penalty on the defendant should be set at $5 million to achieve full deterrent effect.

The question of how the corporate veil doctrine achieves deterrence needs to be analyzed differently for intentional improper conduct and non-intentional, probabilistic improper conduct such as negligence-based torts. This is because as far as intentional conduct is concerned, the perpetrator has control over whether the conduct is undertaken. In order to deter such conduct, the law needs to impose an expected penalty on the perpetrator that is equal to the benefit multiplied by the reciprocal of the probability of apprehension. For probabilistic conduct, the tort may take place despite the precautions undertaken by the tortfeasor. Thus the law needs to ensure that the tortfeasor is liable for the full damages it causes so that it internalizes the full costs of its negligent conduct.[105] In that case, the tortfeasor will take precautions up to the socially efficient level. As the ensuing discussion will show, the corporate veil doctrine is at best only an imperfect tool for achieving deterrence.

The discussion begins with intentional conduct. Shareholder liability resulting from veil piercing falls short of the theoretical requirements of effective deterrence in two important ways. First, the corporate veil doctrine does not impose penalties nor does it directly take into account the shareholder's gains from his improper conduct. The doctrine merely holds shareholders responsible for unpaid corporate liabilities. The outstanding liabilities may have little to do with the perpetrator's ill-gotten gains. If the gains exceed the outstanding liabilities, merely holding the culpable shareholder liable for the latter will not achieve effective deterrence. In some cases, there may be a close correspondence between the ill-gotten gains and the outstanding liabilities. To the extent that the unpaid liabilities are the direct consequence of shareholder misconduct, those liabilities can be used as a gauge of the shareholder's gains. For instance, if a shareholder removes $1 million of corporate assets and the corporation is unable to repay $1 million of debts as a result, these outstanding liabilities would reflect the shareholder's ill-gotten gains. If the outstanding liabilities were then adjusted for the probability of detection, imposition of shareholder liability would be consistent with the theoretical requirements of deterrence.

---

[105] SHAVELL, *supra* note 90, at 236-38.

Second, and more importantly, achieving effective deterrence of intentional conduct requires the unpaid liabilities to be adjusted for the probability of detection. The case law shows that this is not done by the courts. Without such an adjustment, the corporate veil doctrine will not effectively deter intentional shareholder misconduct. One mitigating factor for this deficiency is that the probability of detecting shareholder misconduct is likely to be high given that the consequence of such misconduct is corporate failure and default. There should be little difficulty with detection because the creditors will surely become aware of a default, after which shareholder misconduct will be uncovered. In the event that the shareholder misconduct does not result in a default, detection will not be as straightforward. However, one may argue that the corporate veil doctrine should not be concerned with shareholder misconduct if it does not result in failure to meet corporate liabilities. Deterrence of such conduct is the responsibility of other areas of corporation law. As long as the creditors are fully paid off, the doctrine should have no application.

Deterrence of probabilistic conduct is relatively straightforward. The main type of such conduct as far as the corporate veil doctrine is concerned is negligence-based torts committed by the corporation. As mentioned earlier, to the extent that the underlying tort damages were already set to achieve deterrence, preservation of the deterrent effect only requires that the corporation and its shareholders fully internalize the harm caused by the tort. In other words, the corporate veil doctrine only needs to shift unmet liabilities to the shareholders, which is precisely what the doctrine does. However, the courts have almost never held the shareholders personally liable for a corporate tort absent misconduct on their part. The courts have repeatedly affirmed that the corporate entity can be legitimately employed to limit future tort liabilities.[106] The misconduct required to pierce the veil is always intentional. Absent such misconduct, the full deterrent effect of a tort judgment against a judgment-proof corporation will be curtailed. What would be needed to preserve the full deterrent effect of the judgment is a general rule of unlimited liability.[107] In light of the requirement of intentional misconduct for veil piercing, one may even say that the corporate veil doctrine is never concerned with deterrence against probabilistic torts.

An interesting observation about using the corporate veil doctrine to achieve deterrence is that the deterrent effect is inversely proportional to the defendant's shareholding in the corporation. This can be illustrated with another numerical example. Assume that there are five shareholders in a corporation, only one of whom has dominated the corporation and was found to have committed improper conduct, the detection of which has a probability of 20%. A tort victim files a veil piercing claim against this shareholder, seeking to hold him personally liable for $5 million of unmet liability award. If the claim prevails, personal liability for this shareholder would be $5 million, as opposed to the $1 million which would have been allocated to him had the additional liability been apportioned equally among the five shareholders. Thus the liability multiple is five, which is the reciprocal of his 20%

---

[106] Browning-Ferris Indus. of Ill., Inc. v. Ter Maat, 195 F.3d 953, 960 (7th Cir. 1999) (no special obligation on corporation to capitalize adequately for tort liabilities); Craig v. Lake Asbestos of Que., Ltd., 843 F.2d 145, 150 (3d Cir. 1988) ("[E]vasion of tort liability has never, in itself, been sufficient basis to disregard corporate separateness."); Harris v. Curtis, 87 Cal. Rptr. 614 (Cal. Ct. App.1970).

[107] *See generally* Hansmann & Kraakman II, *supra* note 38, at 1882-94; Leebron, *supra* note 38, at 1584-85.

shareholding.[108] Assuming that the outstanding liability reflects the ill-gotten gains, the shareholder's 20% share ownership inadvertently helps to achieve full deterrent effect. In fact, to the extent that the other shareholders have shared the benefits of the improper conduct, this deterrent effect will be further enhanced. The benefit received by the culpable shareholder will be smaller than the outstanding liability (as it has been shared by other shareholders), while his share of the corporate liability remains unchanged.

Whether this deterrence enhancement effect is desirable is of course a different issue. The smaller a shareholder's ownership interest, the less likely it is that the shareholder has control over corporate affairs and should be held responsible for the corporation's default. In fact, one may argue that the deterrent effect should be made greater the larger a shareholder's ownership stake. The irony is that, as mentioned earlier, veil piercing will achieve no deterrent effect in a single-shareholder corporation. There is no concentration of liability and no dilution of ill-gotten gains. Yet there is no more convincing case for deterrence than a single-shareholder corporation, where the sole shareholder's responsibility for the corporation's conduct is beyond dispute. This is another reason why the corporate veil doctrine is a blunt instrument of deterrence.

Aside from its ineffectiveness as a deterrence tool, the altruistic nature of the corporate veil doctrine means that deterrence is unlikely to be a fitting objective for it. As Kennedy said, altruism calls for sharing and mercy, not punishment and deterrence. Deterrence of undesirable conduct that harms the economic interest of other members of society facilitates self-reliance and private initiatives. It allows individuals to rely on their own effort to pursue their goals without the fear that the fruits of their labor will be compromised. Profiteering through fraud, misrepresentation, and other improper conduct that justifies veil piercing violates the individualistic ideal of self-reliance. Deterrence of such conduct is an essentially individualistic enterprise. The foregoing discussion demonstrates that while deterrence no doubt plays a role in veil piercing, it is ill-suited as the principal objective of the corporate veil doctrine. With punishment, compensation, and deterrence rejected as the principal objective of the corporate veil doctrine, the remaining alternative is the prevention of unjust enrichment.

---

[108] In reality, such a high multiple is highly unlikely because in most of the corporate veil cases, the defendant shareholder owns at least a majority of the corporation's shares. Oftentimes the shareholder owns all or almost all of the shares. Therefore, the liability multiple is rarely greater than two.

4. Prevention of Unjust Enrichment of the Shareholders

Prevention of unjust enrichment of the shareholders is the most apposite principal objective of the corporate veil doctrine. Before an attempt is made to explain the corporate veil cases in terms of unjust enrichment, it is important to explain what the term means. As it turns out, a clear definition of unjust enrichment, and its companion concept of restitution, has eluded scholars for generations since the field was first created by the Restatement of the Law of Restitution in 1937.

Scholars cannot even seem to agree on what constitutes unjust enrichment and restitution. One commentator compares unjust enrichment to contracts and torts and posits that it refers to a range of physical events that give rise to liability.[109] Another asserts that unjust enrichment is the single guiding jurisprudential principle that unifies and underpins the law of restitution.[110] Yet another suggests two alternative characterizations of unjust enrichment: an equitable principle that justifies judicially crafted exceptions to rigid legal rules or a descriptive and organizational principle for the wide range of restitution cases.[111] In fact, unjust enrichment is also a cause of action that allows a plaintiff to recover a benefit it has bestowed upon the defendant, where allowing the defendant to keep that benefit would offend notions of justice.[112] The definition of restitution, especially as it relates to unjust enrichment, is even more controversial. One commentator believes that restitution is exclusively concerned with the rectification of unjust enrichment,[113] while another has attributed both a substantive and a remedial component to restitution.[114] Yet another commentator describes restitution as a goal of the law alongside compensation, punishment, and deterrence.[115] The focus of this article is not to resolve the debate about the proper characterization of unjust enrichment and restitution. It is in fact not necessary to resolve it in order to explain the corporate veil cases in terms of unjust enrichment.[116] As applied to the corporate veil cases, this article treats unjust enrichment as the consequence that may result if separate corporate personality is strictly adhered to, and the prevention of unjust enrichment as a principle that allows judges to create narrow exceptions to limited liability.[117]

---

[109] Peter Birks, *Unjust Enrichment and Wrongful Enrichment*, 79 TEX. L. REV. 1767, 1778 (2001) ("Unjust enrichment has to be understood as an event and, more specifically, a causative event.").

[110] Andrew Kull, *Rationalizing Restitution*, 83 CAL. L. REV. 1191, 1196 (1995) ("My proposition is that the law of restitution be defined exclusively in terms of its core idea, the law of unjust enrichment.").

[111] Emily Sherwin, *Restitution and Equity: An Analysis of the Principle of Unjust Enrichment*, 79 TEX. L. REV. 2083, 2084 (2001). In fact, Professor Sherwin rejects the suggestion that unjust enrichment is a principle that justifies judicially crafted exceptions to rigid legal rules. While her rejection of this characterization of unjust enrichment may be valid for the law of restitution, it will be argued that unjust

---

enrichment does in fact serve as a principle that justifies judicially created exceptions to the rule of limited liability.

[112] *See, e.g.*, Spires v. Hosp. Corp. of Am., 289 F. App'x 269, 272-73 (10th Cir. 2008) ("Under Kansas law, the elements of an unjust enrichment claim are (1) plaintiff conferred a benefit on defendant; (2) defendant knew and received a benefit; and (3) defendant retained the benefit under circumstances that make it unjust.").

[113] Kull, *supra* note 110, at 1196.

[114] Douglas Laycock, *The Scope and Significance of Restitution*, 67 TEX. L. REV. 1277, 1279-80 (1989).

[115] Birks, *supra* note 109, at 1772. Professor Birks' characterization of restitution perhaps suggests that restitution would be a more appropriate term than prevention of unjust enrichment as a description of the goal of the corporate veil doctrine. In his view, restitution, compensation, deterrence, and punishment are comparable concepts. However, given the seemingly endless controversy and misunderstanding surrounding the definition and scope of restitution, this article will use prevention of unjust enrichment instead. Despite disagreements about whether unjust enrichment is a liability-creating event or a principle, there seems to be a relative consensus of what the term means.

[116] The proper characterization of unjust enrichment is mostly pertinent to disentangling the relationship between unjust enrichment and restitution.

[117] Professor Sherwin argues that unjust enrichment should not be treated as a principle allowing judges to create exceptions to clearly defined legal rules which would otherwise apply in restitution cases. She observes that there is nothing inherent in the field of restitution that renders it particularly suitable for individualized decision making. Sherwin, *supra* note 111, at 2096-2101. Her arguments, of course, are

It is important to clarify that by attempting to explain the corporate veil cases in terms of unjust enrichment, this article does not characterize veil piercing as restitution claims.[118] While there are cases in which one can analogize the imposition of shareholder liability to the disgorgement of the benefit received by the defendant, there are cases in which the analogy is inapt.[119] Instead, the focus will be on how veil piercing prevents the unjust enrichment of shareholders that follows from a strict adherence to limited liability.

To demonstrate that unjust enrichment provides a fitting rationale for the corporate veil cases requires a clearer definition of the term. While the concept has great intuitive appeal, even the draft Restatement of Restitution and Unjust Enrichment itself does not seem able to define unjust enrichment in a concrete manner. Section 1 of the draft Restatement asserts that "[a] person who is unjustly enriched at the expense of another is liable in restitution to the other."[120] It has also been said that unjust enrichment is "the equitable principle by which one who has been enriched at the expense of another, whether by mistake, or otherwise, is under a duty to return what he has received or its value to the other."[121] These merely state the legal consequence of an instance of unjust enrichment. The appending comment on Section 1 proceeds to provide a number of examples of unjust enrichment without attempting to offer a formal definition of the term. In a 1938

---

concerned with restitution and do not shed light on the suitability of using prevention of unjust enrichment to explain corporate veil cases. Moreover, it was demonstrated in the previous section that the corporate veil doctrine requires individualized decision making. Therefore, unjust enrichment could be used as a principle allowing judicially crafted exceptions to limited liability so long as the rationale of unjust enrichment fits the corporate veil cases.

[118] Laycock points out that in fact, restitution need not be the only possible remedy for unjust enrichment. There are cases in which damages will be the more appropriate remedy. Laycock, *supra* note 114, at 1285 ("In some cases, unjust enrichment best explains substantive liability, but plaintiff's loss is the most appropriate measure of recovery.").

[119] RESTATEMENT (THIRD) OF THE LAW OF RESTITUTION AND UNJUST ENRICHMENT § 2 (Discussion Draft, March 31, 2000) ("Liability in restitution is based on and measured by the receipt of a benefit, but the receipt of a benefit does not of itself make [the] recipient liable in restitution.").

[120] *Id.* at § 1.

[121] Warren A. Seavey, *Problems in Restitution*, 7 OKLA. L. REV. 257, 257 (1954).

article that explained the then-emerging field of restitution to a skeptical British audience, the reporters of the First Restatement provided a more concrete formulation of the concept. They wrote that "[a] person has a right to have restored to him a benefit gained at his expense by another, if the retention of the benefit by the other would be unjust."[122] This statement at least explains what constitutes enrichment.

As for the meaning of unjust, the comment on Section 1 of the draft Third Restatement suggests that it entails "a direct appeal to standards of equitable and conscientious behavior."[123] This appeal is in many ways reminiscent of the determination of improper conduct under the corporate veil doctrine. There is no single uniform standard for deciding what constitutes improper conduct that justifies veil piercing. The courts make a determination based on notions of fundamental fairness and justice on a case-by-case basis. The draft Restatement similarly highlights the open-ended nature of unjust enrichment by characterizing it as "the embodiment of natural justice and equity [that] gives the subject an undoubted versatility, an adaptability to new situations."[124] The parallel between this characterization of unjust enrichment and the corporate veil doctrine is unmistakable. Integrating these various observations, unjust enrichment can be defined as an equitable principle under which natural justice and equity require the recipient of a benefit to return that benefit to the plaintiff when allowing the recipient to keep it would result in injustice.[125]

How then are shareholders of a corporation unjustly enriched at the expense of the creditors?[126] The benefit conferred depends on the creditor at issue. For a financial creditor, the benefit will be the proceeds of a loan or purchase money for corporate bonds. For a trade creditor, the benefit will be the goods or services supplied to the corporation. For an employee, the benefit will be the labor services rendered. To the extent that the corporation, and by extension the shareholders, receive these benefits without making full payment for them and the default risks have not been fully compensated for, the shareholders can be said to have been enriched. It is obvious how shareholders are enriched if these benefits have not been fully paid for. However, there will be no enrichment if the default risk has been fully compensated. Given adequate compensation, the transaction would simply be a fully bargained-for exchange, and there would be no shareholder enrichment. This means that given that financial creditors are usually in a position to negotiate for a higher rate of return for the default risk, it will be difficult for such a creditor to argue that the shareholders have been unjustly enriched at its expense. Meanwhile, trade creditors and employees, which are in a much weaker position to do the same, should have less difficulty demonstrating enrichment.

---

[122] Warren A. Seavey & Austin W. Scott, *Restitution*, 54 L.Q. REV. 29, 32 (1938).

[123] RESTATEMENT (THIRD), *supra* note 119, at § 1.

[124] *Id.*

[125] Sea-Land Servs., Inc. v. Pepper Source, 993 F.2d 1309 (7th Cir. 1993). In that case, the court defined unjust enrichment as "the receipt of money or its equivalent under circumstances that, in equity and good conscience, suggest that it ought not to be retained because it belongs to someone else." *Id.* at 1312. Based on this definition, the court proceeded to pierce the veil because the shareholder at issue was unjustly enriched.

[126] In contractual corporate veil cases, the contract is between the corporation and the creditor. The creditor, who is trying to recover from the shareholders, does not have a direct contractual relationship with them. This is similar to most unjust enrichment cases in which the plaintiff and the defendant do not share a contractual relationship.

For a tort victim, the main benefit will be the unpaid tort liability that would have compensated the victim. One further benefit conferred by the tort victim on the corporation is reduced accident prevention costs, the incurrence of which would have prevented the tort at issue. As a result of limited liability, the corporation does not fully internalize its accident costs and loses the incentive to undertake adequate precautions.[127] The corporation will lower the level of prevention as a result and achieve savings in prevention costs. The shareholders are enriched to the extent that such savings are shared with them through dividend distribution and other means.

In tort corporate veil cases, there is usually no contractual relationship of any kind between the tort victim and the corporation. Therefore, unjust enrichment applies even more aptly to the tort cases.

[127] Hansmann & Kraakman II, *supra* note 38, at 1882-94.

Mere enrichment is of course different from unjust enrichment.[128] Shareholder enrichment only becomes unjust if the circumstances under which it arises can be considered unjust under notions of natural justice or equity. This means that absent impropriety on the part of the shareholders, mere non-satisfaction of corporate liabilities would not constitute unjust enrichment. The law has made a policy choice that under limited liability, shareholders may be enriched at the outside creditors' expense in order to encourage entrepreneurship and capital flow to corporations. If limited liability is to be upheld, the corporate veil doctrine cannot deem non-payment of corporate liabilities absent shareholder impropriety to be unjust. Otherwise, limited liability would be practically overturned. It is only when the enrichment results under unjust circumstances, such as when the shareholders have misappropriated corporate assets, hence leaving insufficient assets to satisfy the outside creditors, that veil piercing is justified. In this instance, the unjust enrichment equals the amount of assets misappropriated by the shareholders that would have been available to the creditors.

Recasting the corporate veil doctrine in terms of unjust enrichment means that conduct that unjustly enriches shareholders should carry more weight in the analysis. For example, among the seven types of improper conduct enumerated by Powell, fraud, asset stripping, and uncompensated torts are the most likely to result in unjust enrichment. The courts should be more ready to pierce the veil in the presence of such conduct. Unjust enrichment is also relevant to the analysis under the formalities prong of the doctrine. A variety of factors have been identified as indicia of the lack of formal integrity of the corporation. However, not all of them result in unjust enrichment of the shareholders. For example, non-observance of corporate formalities and overlap in corporate personnel are unlikely to contribute directly to shareholder enrichment. These factors should receive less emphasis in the analysis.

---

[128] Laycock points out that not all instances of enrichment are considered unjust under the law of restitution. He gives as examples duress and the conferment of benefit in the absence of mistake, coercion, request, or emergency. Laycock, *supra* note 114, at 1284.

While explaining the corporate veil doctrine in terms of unjust enrichment may seem surprising at first glance, there is substantial support for it. As mentioned earlier, the draft Restatement states that unjust enrichment is premised on notions of natural justice and equity. This echoes the courts' repeated assertions that the corporate veil doctrine is ultimately about preventing injustices. The characterization of unjust enrichment as an equitable principle that allows the courts to create individualized exceptions to general legal rules also aptly describes the role of the corporate veil doctrine as a standard-based exception to the general rule of limited liability. The courts have in fact used unjust enrichment of the shareholders as a basis for veil piercing in some cases.[129] Moreover, prevention of unjust enrichment as the main purpose of contractual remedies has long been affirmed. In his seminal article on contractual damages, Professor Fuller declared that "it would seem that the fundamental purpose of the law is the prevention of unjust enrichment, since enrichment of the promisor at the expense of the promisee is the *sine qua non* of judicial interference."[130] Unjust enrichment of the tortfeasor at the expense of the tort victim is equally deserving of judicial interference, if not even more so.[131] Lastly, attributing prevention of unjust enrichment as the principal objective of the corporate veil doctrine is consistent with its altruistic nature. While deterrence and punishment no doubt play a significant role in some corporate veil cases, they cannot be the principal focus of an altruistic legal doctrine. They have little to do with the primary altruistic concerns of sharing and sacrifice.[132] Meanwhile, unjust enrichment compels the recipient of an ill-gotten benefit to return it to the rightful beneficiary. It requires those who were unduly advantaged by the existing legal regime to share what they not ought to keep. This is in keeping with Kennedy's statement that "[i]f altruism is the sharing and sacrifice of advantages that one might have kept for oneself, then the state forces the strong to behave altruistically."[133] Here, the strong are the shareholders who have unjustly benefited from the protection of limited liability. The corporate veil doctrine compels them to return the benefit to the creditors on grounds of fairness. Unjust enrichment hence reinforces the altruistic tendencies of the doctrine.

---

[129] *Sea-Land Servs.*, 993 F.2d at 1309; Liberty Mut. Ins. Corp. v. M & O Springfield Co., No. 97-4146, 1998 WL 894654, at *4 (7th Cir. Dec. 17, 1998) ("As we previously observed, absent a showing of fraud or unjust enrichment, the fact that there is an unsatisfied judgment is an insufficient ground to compel the extraordinary remedy of piercing the corporate veil."); Tr. of Chi. Painters v. Destiny Decorators, Inc., No. 07-C-4236, 2009 WL 3188687, at *11 (N.D. Ill. Sept. 30, 2009) ("This deprivation, and McDuffie's consequential unjust enrichment, is the sort of fraudulent wrong beyond the creditor's inability to collect sufficient to satisfy the second condition needed to pierce the corporate veil." (internal quotation marks omitted)); Langston v. Rizza Chevrolet, Inc., No. 07-CV-5965, 2008 WL 5083111, at *6 (N.D. Ill. Nov. 24, 2008) (citing unfair enrichment as one ground for veil piercing); United States *ex rel.* Purcell v. MWI Corp., 520 F. Supp. 2d 158, 173 (D.D.C. 2007) ("Here, the defendant is a stockholder of a corporation that received a benefit from the plaintiff. The plaintiff may only rely on an inference that a stockholder by means of his corporate equity received a benefit if the plaintiff shows that the stockholder abused the corporate form, using it as his own alter ego to perpetrate fraud—in which case, the corporate veil should be pierced."); Robertson-Ceco Corp. v. Cornelius, No. 3:03cv475/RV/EMT, 2007 WL 1020326, at *5 n.6 (N.D. Fla. Mar. 30, 2007) ("[U]njust enrichment is relevant to the analysis insofar as it represents one way of determining whether the corporate veil should be pierced."); Munder v. Circle One Condo., Inc., 596 So. 2d 144, 145 (Fla. Dist. Ct. App. 1992) ("Fraud, self-dealing, unjust enrichment and betrayal of trust, may well result in individual [shareholder] liability."); Taylor v. Wellington Station Condo. Ass'n, 633 So. 2d 43, 45 (Fla. Dist. Ct. App. 1994) (stating same).

5. Summing Up

The foregoing analysis of the formal and substantive dichotomies of the corporate veil doctrine has provided a number of important conclusions and observations. First, it demonstrates the relationship between limited liability and the corporate veil doctrine as one between a rule and a standard-based exception to the rule. In particular, it confirms the open-ended nature of the doctrine, which is an essential characteristic of a standard. This means that a certain degree of non-conclusiveness and open-endedness in application is inherent in the doctrine. Effort to enhance its formal realizability beyond a certain point is likely to be futile if not counterproductive. One may argue that this relative lack of formal realizability will hamper the ability of the doctrine to deter undesirable conduct. If the corporate veil cases cannot even inform potential perpetrators in advance precisely what conduct will result in shareholder liability, these perpetrators will find it difficult to adjust their conduct accordingly. This argument, however, should not be overstated for a number of reasons. As has been conclusively argued in this section, the principal objective of the corporate veil doctrine is not deterrence, but prevention of unjust enrichment. If the achievement of the latter entails reduced efficacy in attaining the former, this is a price that has to be paid. Moreover, even though the corporate veil doctrine is a standard of considerable open-endedness, there are some clearly identifiable trends. As Thompson's survey shows, misrepresentation is a highly predictive factor in the veil piercing decisions.[134] Similarly, the courts have repeatedly said that a finding of fraud justifies veil piercing.[135] Some commentators have created lists of do's and don't's for shareholders to avoid veil piercing.[136] There is hence sufficient guidance to potential perpetrators on what conduct to avoid. Lastly, the vagueness of the doctrine may actually enhance its ability to deter by encouraging potential perpetrators to steer clear of the boundary of permissible conduct.[137] In contrast, bright-line rules will allow potential perpetrators to tailor their conduct to the precise scope of the rule. They will make sure that their actions lie just outside the scope of prohibition. What seems at first glance to be a weakness of the doctrine as an instrument of deterrence turns out to be a strength.

---

[130] Lon L. Fuller & William R. Perdue, Jr., *The Reliance Interest in Contract Damages: 1*, 46 YALE L.J. 52, 67 (1936).

[131] As mentioned earlier, many commentators have argued that there are more persuasive arguments for applying unlimited liability to tort victims. Halpern et al., *supra* note 34, at 145-47; Leebron, *supra* note 38, at 1574-87; Hansmann & Kraakman II, *supra* note 38, at 1882-94.

[132] Kennedy, *supra* note 18, at 1717.

[133] *Id.* at 1919.

---

[134] Thompson II, *supra* note 11, at 1063.

[135] Dole Food Co. v. Patrickson, 538 U.S. 468, 475 (2003) ("The doctrine of piercing the corporate veil, however, is the rare exception, applied in the case of fraud or certain other exceptional circumstances . . . ."); United States v. Bestfoods, 524 U.S. 51, 62 (1998) ("[T]he corporate veil may be pierced and the shareholder held liable for the corporation's conduct when, *inter alia*, the corporate form would otherwise be misused to accomplish certain wrongful purposes, most notably fraud, on the shareholder's behalf."); Ill. Bell Tel. Co. v. Global NAPs Ill., Inc., 551 F.3d 587, 597 (7th Cir. 2008) ("What is true is that in a contractual veil-piercing case, such as this case, Delaware permits piercing the veil only upon proof either of fraud or that the corporation simply functioned as a façade for the dominant shareholder."); Dusharm v. Elegant Custom Homes, Inc., 302 F. App'x 571, 572 (9th Cir. 2008); Belvedere Condo. Unit Owners' Ass'n v. R. E. Roark Cos., 617 N.E.2d 1075, 1086 (Ohio 1993).

[136] Krendl & Krendl, *supra* note 11, at 52-55.

[137] Kennedy, *supra* note 18, at 1696.

Second, the analysis shows that the corporate veil doctrine is a hybrid formalities-conduct directive. This should not come as a surprise, given the structure of some of the prevailing formulations of the corporate veil doctrine. Kennedy's assertion about the lawmaker's neutrality toward compliance with formalities affirms the general judicial approach that veil piercing will not be granted merely on the grounds of non-observance of corporate formalities, broadly understood to include factors such as inadequate capitalization and others. Impropriety is needed to substantiate a veil piercing claim. Third, the analysis affirms the doctrine's altruistic underpinning. The doctrine is meant to alleviate the harshness that may result from an unyielding adherence to the general rule of limited liability. Some have argued that application of the doctrine should be confined to exceptional circumstances because veil piercing discourages entrepreneurship and investment in corporations. However, as explained above, the argument that veil piercing undermines the benefits of limited liability has been overstated. The corporate veil doctrine, as it is currently formulated and applied, does not undermine these benefits. In particular, this is due to the fact that veil piercing requires a showing of impropriety or injustice and generally only results in liability for the culpable shareholder. Bona fide entrepreneurs and investors should not be deterred by the doctrine.

Efforts to identify the principal objective of the corporate veil doctrine sheds further light on the application of the doctrine. The discussion about the circumstances under which shareholders are unjustly enriched suggests that veil piercing should be rarely available for financial creditors, who are most likely to be able to negotiate for an adequate compensation for the default risk borne by them. The same is not true for many trade creditors and employees, whose veil piercing claims should thus be adjudged more leniently. As has been argued previously by other commentators, tort victims' inability to negotiate for *ex ante* protection means that they are almost never adequately compensated for the accident costs they are made to bear. Shareholders are thus unjustly enriched at the expense of these victims. This again confirms that the courts should be more willing to pierce the veil in tort cases.

Making prevention of unjust enrichment the principal objective of the doctrine means that the courts should attach greater importance to conduct and breach of formalities that are more likely to result in shareholder enrichment. These include conduct such as asset stripping and fraud and formalities such as commingling of assets and inadequate capitalization. In particular, formalities requirements that are justified on estoppel grounds should be given less emphasis. This is a helpful insight because the courts tend to adopt a totality-of-circumstances approach to the question of whether formal integrity of the corporation has been compromised, without giving much explanation of the relative importance of the different factors. Lastly, if the goal is to prevent unjust enrichment, the corporate veil doctrine should expand the scope of the plaintiff's recovery beyond the unpaid liability to include other forms of enrichment of the shareholder. This last point is probably the most radical departure from the prevailing approaches to the doctrine.

III. REFORMULATION OF THE CORPORATE VEIL DOCTRINE BASED ON AN UNJUST ENRICHMENT RATIONALE

The conclusions drawn above provide a basis upon which the doctrine can be reconceptualized and reprioritized. Many commentators have remarked on the lack of clarity in the doctrine and the difficulty with predicting the outcome of a veil piercing case.[138] The totality-of-circumstances approach adopted by many courts requires them to weigh the myriad factors in an indeterminate manner. The altruistic nature and the inherent open-endedness of the doctrine as a standard mean that some degree of unpredictability is inevitable. Greater clarity nonetheless can be attained with an improved understanding of the formal and substantive nature of the doctrine. The various factors considered by the courts to determine the formal integrity of the corporation are not equally relevant to the veil piercing analysis. In fact, viewed in light of its shareholder unjust enrichment rationale, it will be argued below that most of them do not serve a useful purpose and should be de-emphasized. Our improved understanding of the form and substance of the doctrine will also shed light on the impropriety prong. The guidance provided by the courts and the analysis by the commentators on this prong is scant. What follows is an application of the conclusions from the previous section to fill these gaps and to reformulate the doctrine.

---

[138] Krendl & Krendl, *supra* note 11, at 22 ("We know, however, of no successful effort to integrate all of these issues into a single simple rule which incorporates the proper considerations with sufficient specificity to ensure uniformity and predictability of application."); Thompson II, *supra* note 11, at 1036 ("Piercing the corporate veil is the most litigated issue in corporate law and yet it remains among the least understood."); Barber, *supra* note 8, at 374-75 (providing list of nineteen factors which courts deem relevant for veil piercing); Landers, *supra* note 68, at 620 (describing corporate veil doctrine as "[d]evoid of any consistent doctrinal basis, the cases

A. The Formalities Prong

1. Non-observance of Corporate Formalities

One inherent contradiction in the corporate veil doctrine is that the doctrine determines an issue of substance with reference to form. Liability is an issue of substance. It is about the allocation of economic loss between private parties. Yet much of the analysis under the doctrine focuses on the form of the corporation. This is in some ways understandable because a corporation is a legal fiction. It is an artificial entity that is constituted through a formal process of incorporation. With the abolition of the minimum capital requirement in almost all the states, the incorporation process has been reduced largely to a set of procedures. Once these procedures have been complied with, a corporation comes into being and its separate legal personality is recognized in the eyes of law. However, legal

---

themselves [as] defy[ing] any attempt at rational explanation"); Fischel & Easterbrook, *supra* note 30, at 89 (describing piercing as happening "freakishly" and noting that "the whole area of limited liability, and conversely of piercing the corporate veil, is among the most confusing in corporate law").

form does not determine economic substance. Compliance with the formalities of a corporation does not endow the corporation with economic substance. Conversely, non-observance of those formalities does not deprive the corporation of its economic existence.

Given this disconnect between form and substance, one may legitimately question whether observance of corporate formalities has any place in the corporate veil doctrine. A number of commentators have downplayed the importance of corporate formalities.[139] Similarly, some courts have upheld separate corporate personality despite failure to observe them.[140] In fact, there are instances in which these formalities serve little practical purpose as far as the protection of outside creditors is concerned. The single-shareholder corporation aptly illustrates this. In such a corporation, the requirement of board approval of important corporate decisions is meaningless because the single shareholder will mostly likely control the board. Any proposal from him is assured of approval regardless of the required procedures. Formalistic distinctions do not alter the fact that the sole shareholder has complete substantive control over the corporation. Every corporate act is essentially his act. Therefore, whether this shareholder has complied with corporate formalities should not determine the outcome of the veil piercing analysis.

This is not to say that the form of a corporation does not matter. Corporate procedures and formalities clearly serve important purposes. The purpose they serve often depends on the formality at issue. As Kennedy suggested, formalities directives serve evidentiary and cautionary functions.[141] Some corporate formalities serve a cautionary function by conveying to the parties involved in a particular transaction their intentions to each other. They may also serve an evidentiary function by conveying to the judge what the parties intended with the transaction. In fact, the cautionary function served by formalities in corporation law goes beyond that served by contractual formalities in that they also provide parties not immediately involved in the transaction with useful information. This is perhaps more appropriately called an informational function. This makes intuitive sense because a contract generally only affects the contractual parties. Therefore, the cautionary function need not go beyond them. Meanwhile, the impact of a corporate transaction will be felt by those not immediately involved in it. Corporate claimants such as shareholders and outside creditors may have an interest in the outcome of the transaction, in particular to ensure that their claims against the corporation are not compromised. Some corporate formalities also perform a more substantive function. They ensure that important corporate decisions are well considered and interests of the various corporate claimants are taken into account. This can be called a protective function in the sense that these formalities protect the interests of corporate claimants.

---

[139] Clark, *supra* note 24, at 553; Hamilton, *supra* note 1, at 990-91; Barber, *supra* note 8, at 381.

[140] Zubik v. Zubik, 384 F.2d 267 (3d Cir. 1967); Barlow v. Budge, 127 F.2d 440 (8th Cir. 1942).

[141] Kennedy, *supra* note 18, at 1691.

Whether the observance of corporate formalities has any place in the corporate veil doctrine (outside of single-shareholder corporations) arguably depends on the type of formalities at issue. There seems to be no reason for the doctrine to be concerned with formalities that serve a cautionary or an informational function. The doctrine determines whether shareholder liability should be imposed. Apart from the case where the plaintiff argues that non-compliance with formalities leads to confusion on its part as to whether it is dealing with an individual or a corporation, corporate formalities that serve a cautionary or an informational function should have few implications for shareholder liability. There seems to be a stronger case for formalities that serve a protective function to be incorporated in the doctrine. Observance of these formalities will help to protect the interests of the corporate claimants. Here the problem is that while the doctrine aims to prevent shareholders from unjustly enriching themselves at the expense of outside creditors, many corporate formalities are chiefly designed to protect minority shareholders.[142] It is unclear why failure to accord the protection required by law to minority shareholders should lead to judicial intervention on behalf of the outside creditors. Conversely, compliance with a formality that protects minority shareholders does not mean that creditor interest has been upheld and veil piercing is unwarranted. If the entire body of shareholders decides to siphon off corporate assets for their own personal benefits, a requirement that the removal of assets be approved by a shareholder vote will not protect the creditors. Therefore, even formalities that serve a protective function are not necessarily relevant to the veil piercing analysis.

Even if non-observance of corporate formalities is deemed important enough to warrant judicial intervention, it is not clear that the sanction should be shareholder liability. Kennedy observes that the common legal sanction for non-compliance with a formalities directive is non-enforcement of the transaction.[143] For example, if an important corporate transaction that is required by law to be approved by the board has not received board approval, the law can invalidate the transaction and refuse to enforce it. Perhaps fines can be imposed to encourage compliance. However, if the objective of the sanction is to encourage compliance, shareholder liability seems to be an excessive judicial response, especially when most of these formalities have little to do with harm to the creditors.[144]

The misfit between observance of corporate formalities and veil piercing becomes even more apparent when one considers that estoppel is the most convincing justification that has been put forward for the inclusion of this criterion in the doctrine. As mentioned earlier, the estoppel theory holds that if the shareholders themselves have not treated the corporation as one, they cannot be heard to complain if the court does the same.[145] The validity of this argument, however, depends on the extent of non-compliance. It would be a stretch to say that the shareholders have ignored the existence of the corporation and should be penalized with shareholder liability simply because a shareholder meeting was not convened or a consent was not properly signed. The extent of non-compliance must be extensive for the estoppel argument to hold. A review of the case law suggests that when the courts find that extensive non-compliance justifies veil piercing, it is almost always accompanied by more flagrant shareholder abuses that directly undermine the economic substance of the corporation. An example would be commingling of corporate and personal assets by the shareholders. This inevitably leads one to question the independent significance of non-observance of corporate formalities and wonder whether the courts mention this factor merely to corroborate the lack of economic substance of the corporation.

---

[142] E. R. Latty, *A Conceptualistic Tangle and the One- or Two-Man Corporation*, 34 N.C. L. REV. 471, 474 (1956). Professor Latty argues that corporate formalities, such as the rule that the corporation must act through its board, are often meant to protect the minority shareholders. They do not offer much protection to outside creditors.

[143] Kennedy, *supra* note 18, at 1691.

[144] Hamilton, *supra* note 1, at 990.

The best way to fit non-observance of corporate formalities with the prevention of shareholder unjust enrichment rationale of veil piercing is that observance of these formalities will reduce the likelihood of shareholder abuses and unjust enrichment. While that may be true in some instances, if the concern under the corporate veil doctrine is shareholder unjust enrichment, the analysis should focus on the fact of enrichment itself rather than factors indicating the likelihood of enrichment. These factors are arguably irrelevant if there was in fact no enrichment. The corporate veil doctrine does not intervene merely on the grounds of a heightened probability of shareholder unjust enrichment. It requires actual enrichment. And if there was shareholder unjust enrichment, consideration of these factors would be redundant. There are already criteria in the veil piercing analysis that focus on the fact of enrichment, which means that consideration of non-observance of corporate formalities is largely redundant.

---

[145] Barber, *supra* note 8, at 379.

2. Lack of Substantive Separation

The foregoing discussion suggests that a sound veil piercing analysis should emphasize factors that focus on economic substance rather than form. In light of the principal objective of the doctrine, economic substance should be understood as unjust enrichment of shareholders. Factors that focus on unjust enrichment of shareholders should be given greater weight. Apart from observance of corporate formalities, the courts have considered, under the formalities prong, substantive separation between the shareholders and the corporation, shareholder domination, and overlap of corporate personnel.[146] The ensuing analysis will determine the extent to which these factors are concerned with shareholder unjust enrichment, and hence the weight they should be accorded in the doctrine.

Substantive separation between the shareholders and the corporation refers to the degree of financial independence enjoyed by the corporation. This is determined in light of a range of factors, including whether the corporation maintains its own separate account,[147] whether the salaries of the corporation's employees are paid for by the corporation itself,[148] whether the corporation maintains its own financial records and financial unit,[149] whether the shareholder has advanced loans to the corporation or guaranteed the corporation's loans,[150] and whether the financial transactions between the shareholders and the corporation were well documented and conducted at arm's length.[151]

---

[146] Each of these criteria can be further broken down into more detailed sub-criteria. Many of these sub-criteria are contained on the nineteen-point list developed by Barber and the thirty-one point list developed by the Krendls. Barber, *supra* note 8, at 374-75; Krendl & Krendl, *supra* note 11, at 52-55.

[147] Am. Trading & Prod. Corp. v. Fischbach & Moore, Inc., 311 F. Supp. 412, 414 (N.D. Ill. 1970).

[148] This factor has been emphasized by the courts since the early days of the doctrine. Berkey v. Third Ave. Ry. Co., 155 N.E. 58, 63 (N.Y. 1926) (noting that wages of employees such as motormen and conductors were paid by subsidiary itself).

[149] State v. Swift & Co., 187 S.W.2d 127, 134 (Tex. Civ. App. 1945); *Fischbach & Moore*, 311 F. Supp. at 414. This factor may also be considered under observance of corporate formalities.

[150] Soderberg Adver. Inc. v. Kent-Moore Corp., 524 P.2d 1355, 1362 (Wash. Ct. App. 1974); United States v. Reserve Mining Co., 380 F. Supp. 11, 28 (D. Minn. 1974).

What is most curious and surprising is that some courts have actually held that a shareholder's payment of the corporation's debts may be held against the shareholder and justify veil piercing. In *Ampex Corp. v. Office Electronics, Inc.*, the court cited the parent corporation's continual payment of the subsidiary's debts as a reason for imposing shareholder liability.[152] From a shareholder unjust enrichment perspective, it is not at all clear why shareholder payment of the corporation's debts should result in veil piercing. The corporate parent certainly has not enriched itself by doing so. In fact, it has suffered a financial detriment. One may attempt to defend the court's reasoning on the grounds that the outside creditor may have relied on the fact of shareholder payments when extending credit to the corporation. If that were the argument, however, the plaintiff should be required to provide clear evidence of reliance. In *Ampex Corp.*, there was no reliance on the part of the plaintiff whatsoever because the parent corporation had explicitly informed the plaintiff that it would not take responsibility for the subsidiary's debts.[153] The *Ampex* court's decision would discourage shareholders from shouldering responsibility for corporate debts in the first place. From a shareholder unjust enrichment perspective, there is certainly nothing objectionable about shareholder payment of corporate debts. There is no policy reason to discourage it. If anything, it is beneficial to the outside creditors.

The same argument can be made for shareholder guarantees of corporate debts. Again, the grant of such guarantees in no way results in shareholder unjust enrichment; shareholders in fact suffer a detriment. There is hence no reason to view such guarantees with suspicion. Moreover, shareholders of small close corporations are often required by contractual creditors, especially financial creditors, to provide personal guarantees. If the provision of such guarantees may result in shareholder liability, a narrow carveout from limited liability given to one creditor may effectively result in general shareholder liability. Shareholders would become doubly reluctant to grant personal guarantees, which may impair access to capital by close corporations. The reliance argument should have no application to personal guarantees of debts because a creditor cannot logically assume from the fact that such a guarantee has been given to another creditor, or the same creditor for a previous debt for that matter, that the shareholder will bear personal liability for other debts. From a shareholder unjust enrichment perspective, shareholder payment of the salaries of the corporation's employees is no different from shareholder payment of corporate debts or grant of personal guarantees. Again, there is no reason that payment of salaries should result in shareholder liability. These three considerations should be left out of the determination of the financial independence of the corporation, and should not be considered in the veil piercing analysis.

---

[151] *Fischbach & Moore*, 311 F. Supp. at 414.

[152] 320 N.E.2d 486, 489 (Ill. App. Ct. 1974).

[153] *Id.* at 23.

As for the remaining factors of whether the corporation maintains its own separate account, whether the corporation maintains its own financial records and financial unit, and whether the financial transactions between the shareholders and the corporation were well documented and conducted at arm's length, they tend to focus on formalities and structure that guard against shareholder abuses. In other words, similar to observance of corporate formalities, they do not go to the substantive issue of shareholder unjust enrichment, but only to the likelihood of enrichment. If a shareholder has misappropriated corporate assets for personal use, it should matter little for the purpose of shareholder liability whether the corporation has in place a state-of-the-art accounting system or whether other legitimate corporate transactions were well documented. In short, lack of substantive separation between the corporation and the shareholders should be pared down to shareholder misappropriation of corporate assets. The remaining considerations do not focus on shareholder unjust enrichment and are not helpful to the veil piercing analysis.

3. Shareholder Domination and Overlap in Corporate Personnel

Shareholder domination refers to excessive shareholder control of the corporation. Excessiveness can be understood as going beyond what is acceptable commercial practice. From early on, the courts have held that shareholder domination must be more than complete share ownership.[154] This is to be expected because to hold otherwise would amount to a wholesale repeal of limited liability between a corporation and its wholly owned subsidiaries, which the courts are understandably reluctant to do without legislative intervention.[155] There is, however, no clear threshold for excessive control. No courts have articulated how much more control than complete share ownership would suffice to establish domination. Instead, the courts have come up with vague formulations such as "[w]hen a parent corporation dominates a subsidiary so that the subsidiary is an instrumentality or agent of the parent corporation"[156] or "a domination and control so complete that the corporation may be said to have no will, mind, or existence of its own."[157] These formulations are long on metaphors but short on clarity. Instrumentality and agency are labels rather than operative standards and cannot be used to decide concrete cases.[158] Given that a corporation is an inanimate entity, it cannot have a will or mind of its own, and its existence is derived from its legal recognition. The reference to a corporation's existence in the determination of shareholder domination is circular reasoning, and the reference to will and mind may qualify as what Professor Felix Cohen would call "transcendental nonsense."[159]

---

[154] Craig v. Lake Asbestos of Que., Ltd., 843 F.2d 145, 150 (3d Cir. 1988) ("The control which a parent must exercise over a subsidiary so as to warrant piercing the veil between them is more than 'mere majority or complete stock control' . . . ."); Lowendahl v. Balt. & Ohio R.R. Co., 287 N.Y.S. 62, 72-73 (N.Y. App. Div. 1936) ("Control through mere ownership of a majority or of even all the capital stock and the use of the power incident thereto to elect officers and directors will not in and of itself predicate liability.").

[155] Some commentators have in fact argued for this. *See generally* Landers, *supra* note 68. Regardless of the theoretical merit of this view, the sweeping scope of such a repeal means that it should not be implemented judicially. It should be left to the state legislatures to decide.

[156] Fitzpatrick v. Commonwealth Oil Co., 285 F.2d 726, 730 (5th Cir. 1960).

[157] *Lowendahl*, 287 N.Y.S. at 73.

[158] Thompson II, supra note 11, at 1037 ("A common refrain in the literature is an attack on the use of conclusory terms, such as 'alter ego' and 'instrumentality,' that provide no insight into the nature of the factors considered."). These terms merely state the courts' conclusions rather than provide reasons for the result.

The *Lake Asbestos* court came up with a more concrete formulation for the threshold, stating that domination amounts to shareholder control "not only of finances but of policy and business practice in respect to the transaction attacked so that the corporate entity . . . had at the time no separate mind, will or existence of its own."[160] Despite the inevitable reference to the lack of a separate mind or existence, the *Lake Asbestos* court did provide two useful focal points for the analysis. While control of finances can be dealt with under the rubric of lack of substantive separation and has been discussed in the previous section, control of policy and business practice may provide a fruitful line of inquiry. The question is how much control over policy and business practice must there be for the courts to deem it excessive and hence support a finding of shareholder domination. Some courts have suggested that to qualify as excessive, control must extend to the daily operations of the corporation.[161] However, it is unclear why day-to-day control over the corporation's affairs should raise greater judicial concerns. In a single-shareholder corporation, the shareholder is for all intents and purposes the corporation itself. He controls every aspect of it, from the election of directors to daily operations. There is no reason to attach particular importance to control over daily operation. The conclusions are largely the same for close corporations with individual shareholders. In most closely held corporations, the shareholders, at least the active ones, exercise extensive control over the firm's operations. They may at the same time function as directors and officers. They make important corporate decisions as directors and decide daily operational issues as officers. Again, the control of daily operations should not be accorded independent significance.

---

[159] Felix S. Cohen, *Transcendental Nonsense and the Functional Approach*, 35 COLUM. L. REV. 809 (1935).

[160] Craig v. Lake Asbestos of Que., Ltd., 843 F.2d 145, 150 (3d. Cir. 1988).

[161] *Id.* at 152 ("The involvement in Cape's [subsidiary] financial and managerial affairs fails to rise to the high standard of domination necessary to pierce the corporate veil . . . . In this case, there is no evidence that Charter's [corporate parent] intrusion into Cape's affairs is even 'constant' or day-to-day."); State v. Swift & Co., 187 S.W.2d 127, 134 (Tex. Civ. App. 1945).

The situation is slightly different for a corporate parent and its subsidiary. The parent as a shareholder owns a majority, or perhaps the entirety, of the subsidiary's shares, but it does not run the subsidiary itself. It must do so through directors and officers whom it may appoint. A parent may be deemed to have exercised excessive control over a subsidiary by appointing its own employees or directors as the subsidiary's directors and officers, who then manage the subsidiary's daily operations on the parent's behalf. This is what the courts have called overlap in corporate personnel. In this sense, overlap in corporate personnel is merely a manifestation of shareholder domination in the parent-subsidiary context.[162] The fact that the corporate parent can and did nominate its own personnel as the subsidiary's directors and officers is evidence that the parent exercised a high degree of control over the subsidiary. However, one wonders why such appointments are objectionable. After all, the power to appoint directors is a right that inheres in share ownership. There is generally no restriction on what kind of directors a shareholder may appoint. By piercing the veil on the grounds of overlap in corporate personnel, the courts are implicitly requiring a corporate parent to appoint individuals not under its direct control to run its subsidiary. In other words, the courts are mandating independent directors and officers through the corporate veil doctrine.

This requirement does not serve any immediate purpose from a shareholder unjust enrichment perspective. It would only help to prevent enrichment if the number of independent directors and officers required is sufficiently high to allow these directors and officers to make a difference in the corporate decision making. The courts by and large have not articulated the extent of overlap of corporate personnel necessary to trigger veil piercing. In most of the corporate veil cases, the overlap is rather complete. This suggests that a less-than-complete overlap would most probably not result in shareholder liability. Overlap in corporate personnel as a veil piercing criterion would amount to a mere formality that the parent appoint a handful of non-employees as directors and officers. This would be a lamentable case of form over substance and would be inconsistent with the focus of the corporate veil doctrine on economic substance.

---

[162] Overlap in corporate personnel, as a criterion, is only relevant if the shareholder is a corporation. Only a parent corporation can share the same officers with its subsidiary. An individual shareholder obviously does not have officers of his own.

It is not even clear that shareholder domination and overlap in corporate personnel advance the prevention of shareholder unjust enrichment rationale of the doctrine in any significant way. Shareholder domination and overlap in corporate personnel on their own do not confer financial benefit on the shareholders. A shareholder may own all the shares of a corporation and may have chosen every single board member and officer. It may even require every corporate decision to be personally approved by itself. Again, all this only indicates the likelihood, but not the fact, of shareholder unjust enrichment. The importance of these two criteria should thus be downplayed.

This conclusion can be illustrated by some cases that highlight the interplay between these two criteria and shareholder unjust enrichment. In *Consolidated Sun Ray, Inc. v. Oppenstein*, the parent corporation incorporated a subsidiary for the purpose of leasing a premise for a retail business.[163] The subsidiary was practically an empty shell created for the sole purpose of holding the lease. The court pierced the veil of the subsidiary and held the parent liable for the rent owed by the subsidiary. Despite its remarks on the complete overlap in the directors and officers of the two corporations and the degree of shareholder domination,[164] it was obvious that what the court found objectionable was the blatant attempt to insulate the parent from liability through the use of the subsidiary. After the parent had decided to close down the retail business, the parent made no attempt to settle the outstanding rent, even though the retail business received funds from the sale of its inventory.[165] By exploiting the corporate entity, the corporate parent was unjustly enriched at the expense of the landlord through the avoidance of the rent payment.

---

[163] 335 F.2d 801 (8th Cir. 1964).

[164] *Id.* at 804.

In *Luckenbach S.S. Co. v. W.R. Grace & Co.*, one of the well-known early corporate veil cases, the Luckenbach Company incorporated a subsidiary, the Luckenbach Steamship Company, for the purpose of entering into freight contracts with counterparties.[166] The parent owned all of the ships in the group, which were leased to the subsidiary to allow it to carry out the freight contracts. The subsidiary was in breach of one particular freight contract and the counterparty sued the parent corporation to recover the damages, which the subsidiary did not have sufficient assets to cover. The court pierced the veil of the subsidiary to hold the parent liable, emphasizing, among other things, the fact that the two corporations had the same directors and officers and were controlled by the same shareholders. Again, despite references to shareholder domination and overlap in corporate personnel, the court's ultimate reason for piercing the veil was shareholder unjust enrichment. The court declared that "it would be unconscionable to allow the owner of this fleet of steamers, worth millions of dollars, to escape liability because it had turn them over a year before to a $10,000 corporation."[167] The corporate parent was unjustly enriched in the amount of the contractual damages it would have been able to avoid under limited liability.

More cases can be provided to illustrate the basic point that criteria such as shareholder domination, overlap in corporate personnel, and observance of corporate formalities do not focus on shareholder unjust enrichment—the paramount concern of the corporate veil doctrine—and could be done away with by focusing on the fact of enrichment itself. Their usefulness as criteria is further undermined by their highly malleable nature. The lack of clear thresholds for determining the degree of shareholder domination, overlap in corporate personnel, or non-observance of corporate formalities that justifies veil piercing means that the courts sometimes play fast and loose with them to suit the result they want to achieve.

---

[165] *Id.* at 805.

[166] 267 F. 676 (4th Cir. 1920).

[167] *Id.* at 681.

The famous early corporate veil case of *Berkey v. Third Avenue Railway* aptly illustrates this.[168] It was apparent early in the opinion that Justice Cardozo, then sitting on the New York Court of Appeals, was minded to uphold the separate corporate personality of the subsidiary, probably because a contrary holding would have resulted in criminal liability for the corporations involved.[169] There was evidence of overlap in corporate personnel, shareholder domination, and lack of substantive separation in the case. In deflecting this evidence, Justice Cardozo focused on a few details of minor importance. First, he pointed out that the boards of the two corporations "were nearly, though not quite the same."[170] It is difficult to appreciate the importance of this difference given that they shared the same "president, treasurer, general manager, paymaster and counsel."[171] Second, even though the salaries of the executive officers were paid with temporary advances from the parent corporation, wages for the motormen and conductors were covered by the subsidiary's own funds.[172] Again, there seems to be no principled basis to distinguish the relative importance of the payment of executive salaries as opposed to workmen's wages for the purpose of veil piercing.

In *Zubik v. Zubik*, the court refused to impose liability on the controlling shareholder of the corporation despite copious evidence of lack of substantive separation and non-observance of corporate formalities.[173] The case involved the imposition of shareholder liability for a corporate tort resulting from an accident in a river. The bulk of the assets of the corporation were owned by the shareholder who leased them back to the corporation through very informal leasing arrangements that suggested “a possible lack of arm’s-length dealing.”[174] The court noted that the lack of formalities extended to other important aspects of the corporation’s operations, such as meetings, expenses, and oral renewal of leases.[175] However, the court dismissed the relevance of all this evidence by arguing that “[a] person who travels on the river, like a person who travels on the highway, does not evaluate the financial responsibility or structure of each person who may collide with him.”[176] While that is no doubt a valid observation, one can say the same about contractual creditors, most of whom do not check whether the corporation has been holding regular board and shareholders’ meetings or whether the employees’ salaries are paid by the corporation as opposed to shareholder funds prior to entering into a contract with the corporation. What ultimately drove the *Zubik* court’s decision was not the degree of observance of corporate formalities or substantive separation between the controlling shareholder and the corporation, but the fact that the shareholder was an “old (71 at the time of trial), illiterate, [and] ill man.”[177]

---

[168] 155 N.E. 58 (N.Y. 1926).

[169] *Id.* at 91 ("We feel assured that no such inference is to be drawn from acts so uncertain in their suggestions, where the inference is also one of the commission of a crime.").

[170] *Id.* at 88.

[171] *Id.*

[172] *Id.* at 89.

[173] 384 F.2d 267 (3d Cir. 1967).

It should be obvious from these two cases that absent complete identity between two corporations, complete disregard of corporate formalities or other extreme circumstances, judges can probably draw contrary conclusions from the same facts applying the same set of criteria of substantive separation, shareholder domination, observance of corporate formalities, and overlap in corporate personnel. These criteria not only have failed to provide clarity to the analysis, they can be said to have obfuscated it. The corporate veil doctrine would be better off if it simply refocused on the ultimate issue of shareholder unjust enrichment.

B. The Impropriety Prong

In contrast to the formalities prong, the impropriety prong of the corporate veil doctrine has received relatively less attention from commentators. This lack of attention is both unfortunate and understandable. It is unfortunate because, as the previous section showed, the formalities prong in most instances only focuses on the manifestations rather than the crux of the issue in corporate veil cases. This lack of attention is also understandable in light of the courts' open-ended formulations of what is required under the impropriety prong. The courts have often said that limited liability would not be upheld if doing so would produce an inequitable result,[178] sanction a fraud,[179] or promote injustice.[180] The courts have not articulated clear standards for determining what constitutes an inequitable result, a fraud, or injustice in the veil piercing context. These formulations are left intentionally vague to capture the wide variety of facts that arise in corporate veil cases. But this factual diversity has also rendered it very difficult to undertake a systematic categorization or analysis.

---

[174] *Id.* at 271.

[175] *Id.*

[176] *Id.* at 274 n.17.

[177] *Id.* at 274.

One exception from the lack of scholarly attention in the recent literature is Krendl & Krendl, who listed the following five categories of conduct as satisfying the impropriety prong: (1) violations of public policy, (2) misrepresentation, (3) lack of economic substance,[181] (4) shareholder domination or direct participation in the wrongful conduct, and (5) evasion of legal obligations through the use of a corporate entity.[182] In his original formulation of the instrumentality doctrine, Powell enumerated seven instances of impropriety.[183] What is striking, but also unsurprising about these categories is that many of them explicitly or tacitly focus on shareholder unjust enrichment. Among Powell's categories, fraud, asset stripping, torts, and possibly other cases of wrong or injustice all exhibit such a focus.[184] How torts unjustly enrich shareholders was explained in Section 0. Shareholder unjust enrichment clearly follows from fraud and asset stripping, through which shareholders remove corporate assets for their personal benefit. Among Krendls' categories, lack of economic substance and evasion of legal obligations are also premised on shareholder unjust enrichment. The discussion below will focus on inadequate capitalization, asset stripping, and deliberate unprofitable operation. It will be demonstrated that they are all best explained as instances of shareholder unjust enrichment.

---

[178] Williamson v. Recovery Ltd. P'ship, 542 F.3d 43, 53 (2d Cir. 2008).

[179] Laborers' Pension Fund v. Lay-Com, Inc., 580 F.3d 602, 610 (7th Cir. 2009).

[180] *Id.*

[181] Krendl & Krendl further divided this into deliberate unprofitable operation and under-capitalization. Deliberate unprofitable operation refers to a situation in which the corporation, usually a subsidiary, was set up in such a way that it was never given the opportunity to be viable on its own. Krendl & Krendl, *supra* note 11, at 36-38. These corporations are usually set up with the sole purpose of supplying the parent corporation with raw materials or other inputs of production. United States v. Reserve Mining Co., 380 F. Supp. 11 (D. Minn. 1971); Int'l Tel. & Tel., Corp. v. Holton, 247 F.2d 178 (4th Cir. 1957).

[182] Krendl & Krendl, *supra* note 11, at 28-29.

[183] The seven instances are (1) actual fraud; (2) violation of a statute; (3) stripping the subsidiary of its assets; (4) misrepresentation; (5) estoppel; (6) torts; and (7) other cases of wrong or injustice. F. POWELL, PARENT AND SUBSIDIARY CORPORATIONS, §13 (1931).

1. Inadequate Capitalization

Among all the grounds for veil piercing, inadequate capitalization has probably received the most scholarly attention.[185] Some commentators have called it one of the most important grounds for veil piercing.[186] This attention notwithstanding, there remains much confusion about the scope and the application of the factor. A number of commentators have remarked on the failure on the part of the courts to articulate a clear standard for determining adequacy of capitalization.[187] There is not even an academic consensus on whether inadequate capitalization should result in shareholder liability.[188] Beyond issues of measurement, some commentators have posited that inadequate capitalization should be accorded varying weights in different cases. Their view is that inadequate capitalization is a more important consideration in tort cases than in contract cases.[189] In fact, some courts have argued that in tort cases, inadequate capitalization is more than mere indicia for the formal integrity of the corporation. It is a proxy for improper motive.[190] This reflects confusion regarding the proper place for inadequate capitalization in the doctrine. While the general view seems to be that inadequate capitalization pertains to the formal integrity of the corporation,[191] some commentators have characterized it as an instance of improper conduct.[192] It will be argued below that, properly understood, inadequate capitalization should be considered under the impropriety prong.

---

[184] Violation of a statute under Powell's categorization and violation of public policy under Krendls' classification both mainly refer to circumvention of statutory prohibitions. Neither is concerned with contractual or tort corporate veil cases, which are the focus of this article. Statutory corporate veil cases implicate the policy rationale of the statute at issue and do not concern the prevention of unjust enrichment rationale of the corporate veil doctrine. The rationale of misrepresentation is not shareholder enrichment, but is premised on reliance by the outside creditor on the shareholders' representation. Veil piercing under misrepresentation is justified on either of these two lines of reasoning: first, that the shareholder who has made a representation to the creditor is estopped from repudiating it, and second, given the shareholder representation, the creditor can be deemed to have entered into the contract directly with the shareholders. Black & White, Inc. v. Love, 367 S.W.2d 427 (Ark. 1963); Zaist v. Olson, 227 A.2d 552 (Conn. 1967).

[185] William P. Hackney & Tracey G. Benson, *Shareholder Liability for Inadequate Capital*, 43 U. PITT. L. REV. 837, 854-60 (1981); Leebron, *supra* note 38, at 1613, 1634; Barber, *supra* note 8, at 391-96; Hamilton, *supra* note 1, at 985-89; Douglas & Shanks, *supra* note 75, at 214; Clark, *supra* note 24, at 535-36; Easterbrook & Fischel, *supra* note 34, at 113; Bernard F. Cataldo, *Limited Liability with One-Man Companies and Subsidiary Corporations*, 18 LAW & CONTEMP. PROBS. 473, 484 (1953).

[186] Inadequate capitalization was determinative in many early corporate veil cases. *See* Douglas & Shanks, *supra* note 75, at 214. It continues to be viewed as an important ground for veil piercing. Leebron, *supra* note 38, at 1634. However, the results from Professor Thompson's survey show that inadequate capitalization is not as outcome-predictive as commentators believe. Thompson II, *supra* note 11, at 1065-67.

---

[187] Barber, *supra* note 8, at 391 ("No general rule has been articulated by the courts for determining adequate capital."); Hackney & Benson, *supra* note 182, at 890 ("In virtually no cases have the courts clearly enunciated what they mean by inadequate capital when it is said to be a major factor in the imposition of shareholder liability."); Hamilton, *supra* note 1, at 1001 ("[T]here is no easy way to determine when a corporation is in fact undercapitalized . . . ."); Roger E. Meiners, James S. Mofsky & Robert D. Tollison, *Piercing the Veil of Limited Liability*, 4 DEL. J. CORP. L. 351, 354 (1979) ("[N]o court has formulated a set of concrete criteria for determining adequate capitalization . . . .").

[188] *See* Clark, *supra* note 24, at 535-36 (inadequate capitalization without deception does not compromise creditor interest so long as creditors have opportunity to negotiate for adequate compensation). *But see* Easterbrook & Fischel, *supra* note 34, at 113 (veil piercing based on inadequate capitalization will encourage firms to divulge unusually low capitalization to potential creditors).

[189] Hamilton, *supra* note 1, at 988.

[190] Radaszewski v. Telecom Corp., 981 F.2d 305, 308 (8th Cir. 1992) ("The reason, we think, is not because undercapitalization, in and of itself, is unlawful (though it may be for some purposes), but rather because the creation of an undercapitalized subsidiary justifies an inference that the parent is either deliberately or recklessly creating a business that will not be able to pay its bills or satisfy judgments against it.").

[191] Barber, *supra* note 8, at 386-90.

[192] Krendl & Krendl, *supra* note 11, at 37-38.

The ensuing discussion will address the following issues raised by the commentators and the courts concerning inadequate capitalization: first, whether inadequate capitalization is a relevant consideration under the corporate veil doctrine at all; second, what is the appropriate standard for determining adequacy of capitalization; and third, whether inadequate capitalization should be accorded different treatment and weight in contract and tort cases.

(i) Whether Inadequate Capitalization Should be a Relevant Factor

The first and foremost issue to be addressed is whether inadequate capitalization should be a relevant consideration in veil piercing cases at all. One forceful objection to inadequate capitalization as a veil piercing criterion is based on the incongruity between the wrong and the remedy. It has been argued that if the wrong is inadequate capitalization, the shareholders should only be held liable for the shortfall in capitalization as opposed to the full share of the outstanding liability.[193] This can be illustrated by a numerical example. Assume that a court would consider $40,000 to be adequate capital for a business of a certain size and risk profile, and the actual initial capitalization was $20,000. The liability at issue in the case is $200,000 owed to a trade creditor, who is the only creditor of the corporation. With its capital base, the corporation only manages to pay off $20,000 of the liability. The creditor attempts to hold the shareholders liable for the unpaid liability of $180,000. Should the court pierce the veil for a shortfall of $20,000 in capitalization, the shareholders end up bearing a liability that is nine times larger. If the shareholders had adequately capitalized the corporation, and there were no other independent grounds for veil piercing, the creditor would have received $40,000 as opposed to the full amount of

---

[193] Philippe M. Salomon, *Limited Limited Liability: A Definitive Judicial Standard for the Inadequate Capitalization Problem*, 47 TEMP. L.Q. 321 (1974).

$200,000 (assuming that the shareholders are not judgment proof).[194]

The inclusion of inadequate capitalization as a veil piercing criterion is fully consistent with the shareholder unjust enrichment rationale of the corporate veil doctrine. By inadequately capitalizing the corporation, the shareholders have unjustly enriched themselves by the amount of the capital shortfall, which would have been available to satisfy the outside creditors. If the shareholders had invested an adequate amount of capital in the corporation, the creditors would have had more assets to satisfy their liabilities. This of course does not address the incongruity objection. In fact, this objection is not confined to inadequate capitalization. It could be equally made against shareholder fraud and misappropriation of corporate assets. One may similarly contend that if the shareholders have deprived the creditors of funds available to satisfy their debts in the amount of the misappropriated assets, the shareholders should not be made fully liable for the unmet liability but should only be made to return the misappropriated assets. After all, to the extent that the outstanding liability exceeds the pilfered assets, imposing shareholder liability increases the amount of assets available for creditor recovery beyond what would have been available absent misappropriation. These related objections will be addressed in the remedies section below. For now, suffice it to say that plaintiff recovery under an unjust enrichment claim may exceed the defendant's gain. Therefore, it would be consistent with the doctrine's shareholder unjust enrichment rationale to allow a veil piercing plaintiff to recover more than what the defendant shareholder has gained from its improper conduct.

---

[194] Another example would be the imposition of unlimited shareholder liability for failure to observe corporate formalities. Assume that a corporate loan from a bank was not properly authorized by the board because the controlling shareholder and director of a corporation has not convened a board meeting for five years. Moreover, the shareholder has failed to keep proper corporate records. The business has failed because the shareholder has misjudged the market, and the corporation is now unable to repay the loan. The bank attempts to pierce the corporate veil because the loan was not properly authorized by the board. If the court pierces the veil for the shareholder's failure to observe corporate formalities, one may similarly think that the remedy is punitive in the sense that it far exceeds the harm caused by the shareholder's failure. In fact, given that the business failure and the consequent inability to repay the loan has nothing to do with the absence of board approval, one may argue that the non-observance of corporate formalities has caused the bank no harm.

Dean Clark asserted that at least with respect to contractual creditors, inadequate capitalization is only unfair and warrants legal intervention if the corporation did not disclose the fact and the creditor did not have the opportunity to negotiate for adequate compensation.[195] Similarly focusing on the importance of disclosure but reaching a contrary conclusion, Professors Easterbrook and Fischel argued that veil piercing based on inadequate capitalization will encourage firms to divulge an unusually low level of capitalization to creditors.[196] While Dean Clark's argument seems to question the relevance of inadequate capitalization for veil piercing, his argument is in fact consistent with the shareholder unjust enrichment rationale of the doctrine. The emphasis on prior opportunity to negotiate highlights the importance of whether the creditor has been adequately compensated for the heightened default risk resulting from inadequate capitalization. As suggested earlier, shareholders are only unjustly enriched at the expense of outside creditors if the latter has not been adequately compensated for the default risk. The lack of a prior opportunity to negotiate suggests shareholder unjust enrichment.

Where the shareholder unjust enrichment analysis diverges from Dean Clark's argument is the extent to which one can assume that disclosure to creditors necessarily permits negotiation for a higher compensation. This assumption most likely holds for a financial creditor, which means that veil piercing on the grounds of inadequate capitalization should be rarely made available to a financial creditor who received full disclosure in advance. However, this assumption is unlikely to be valid for employees and for at least some trade creditors. This means that even with *ex ante* disclosure, inadequate capitalization should still be considered in veil piercing claims brought by trade creditors and employees. At the very least, the courts should undertake a detailed inquiry into whether the employee or trade creditor at issue had the opportunity or bargaining power to negotiate for higher compensation before deciding the weight that should be attached to inadequate capitalization. This kind of circumstance-specific inquiry can be easily accommodated in the corporate veil doctrine.

---

[195] Clark, *supra* note 24, at 535-36. Although this point was made in the context of a discussion about equitable subordination, Dean Clark explicitly analogizes equitable subordination to veil piercing. He observes that "equitable subordination is a functional substitute for, though not an equivalent of, a trustee action to 'pierce the corporate veil' under state law." *Id.* at 535. Therefore, his assertion about the unfairness of inadequate capitalization to creditors presumably also applies in the veil piercing context.

[196] Easterbrook & Fischel, *supra* note 34, at 113. Their argument is clearly premised on a deterrence rationale for the corporate veil doctrine, which has been rejected already. Moreover, this deterrence-based argument would only be valid if there is a clear benchmark for inadequate capitalization, which, as will be suggested below, is sorely lacking.

(ii) The Appropriate Test for Capital Adequacy

The issue of the appropriate test for capital adequacy has long befuddled the courts. Commentators have remarked on this judicial inability to articulate a clear-cut standard for capital adequacy.[197] Yet they have been unable to offer anything more than vague formulations such as "capital sufficient to meet the reasonable requirements of the business in question"[198] or "an amount of equity that is within the ordinary range for the business in question."[199] One slightly more concrete suggestion is to compare the level of capitalization of the corporation at issue with other corporations with a similar size and scope of business.[200] This comparative approach, however, is fraught with difficulty. To begin with, it is not always easy to locate suitable comparators for the kind of close corporations that feature in corporate veil cases.[201] Even if these firms can be located, the court must know that the level of capitalization of the comparator firm is itself adequate before a valid comparison can be made. That entails comparing that comparator firm with yet another firm.

---

[197] *See supra* text accompanying note 184.

[198] Meiners et al., *supra* note 184, at 354.

[199] Easterbrook & Fischel, *supra* note 34, at 113.

[200] Barber, *supra* note 8, at 392; Cataldo, *supra* note 182, at 496 n.90. *See In re* Lumber Inc., 124 F. Supp. 302 (D. Or. 1954) (court determined adequacy of capitalization with reference to industry norm).

[201] Recall Professor Thompson's finding that no public corporation has had its veil pierced. Thompson II, *supra* note 11, at 1055.

This iterative process, which presumably would have to go on *ad infinitum*, can be avoided if the court could somehow ascertain the average level of capitalization of all the comparable firms. Yet it is not immediately obvious how any court can undertake this potentially enormous undertaking. Even if this could be done, there is no guarantee that the result obtained represents adequate capitalization. If a majority of the comparable firms in the market are inadequately capitalized, the result of this survey will likely yield a deficient result.

According to Professor Barber, the starting point for most courts in formulating a standard for adequate capitalization is the idea that "capitalization must be adequate to cover the reasonably foreseeable risks of the business."[202] Dean Clark has attempted to put this idea in more concrete application by adapting the famous B<PL formula proposed by Judge Learned Hand in *Carroll Towing*.[203] Under this modified B<PL formula, the veil would be pierced if PL is greater than B, or in other words, if the probability-adjusted loss to the creditor is greater than the cost of an additional increment of equity capital.[204] One fundamental misconception behind this modified B<PL formula is that the provision of additional shareholder capital will materially reduce the probability of default, or to put it differently, that there will be a significant effect on the default risk after a change in capitalization. In most corporations, the equity capital base is insufficient to cover outstanding liabilities. The equity capital of a corporation was never meant to be large enough to cover all liabilities. It was only meant to

---

[202] Barber, *supra* note 8, at 390.

[203] Clark, *supra* note 24, at 545 n.107. According to Dean Clark's adaptation of the B<PL formula to the determination of capital adequacy,

> B can represent the discounted present value of the cost of an additional increment of equity capital; L, the losses that would be suffered by contract and tort creditors if the corporation were to fail; P, that portion of the probability of failure that would be eliminated by supplying the additional increment of equity capital to the corporation; and PL would be discounted to present value.

Id. at 546.

[204] Professor Leebron has criticized this test, asserting that "this approach illustrates rather than cures the problems with the adequate capital test." Leebron, *supra* note 38, at 1634 n.208.

provide a working capital for the corporation and to offer some minimum level of protection to the creditors. In fact, a requirement that the stated capital be sufficient to cover all liabilities would be tantamount to a repeal of limited liability. Instead, liabilities are usually paid off with the revenue or assets generated by the corporation's operation. And it is the risk that the generated assets fall short of liabilities as they come due that causes a corporation to default and fail. This risk, which is what ultimately concerns creditors, will depend on a variety of factors such as the conditions of the market in which the firm operates, fluctuations in the costs of production, or even weather. A small increase in capitalization is unlikely to have significant impact on these risks and will only marginally increase the recovery available to creditors in the event of a corporate failure. The implication is that an increment of additional capital will have insignificant impact on the probability of default, which means that PL is likely to be small and rarely greater than B. Veil piercing hence will be rare.

Given that equity capital is not intended to provide full or even substantial coverage for corporate liabilities, attempts to formulate a capital adequacy test based on a rigorous quantitative relationship between this capital and the corporation's liability needs is likely to be futile. Moreover, so long as the realized liability in any kind of quantitative test is discounted by changes in the probability of default, the test will only call for an increase in risk-adjusted shareholder capital.[205] This risk-adjusted capital will never be sufficient to cover the realized ability, which of course is not discounted by risk. The fundamental reason for this disconnect is that shareholder capital is not insurance for corporate liabilities.[206] The "payout" from shareholder capital in the event of default is the same as the "insurance premium" put in by the shareholders. The kind of quantitative test suggested by Dean Clark is more appropriately utilized for determining the adequacy of insurance coverage, should the law impose a duty to obtain adequate liability insurance on corporations.[207] It is, however, not suitable for determining capital adequacy.

---

[205] *Id.* at 1634 n.208.

[206] For an interesting discussion of the interplay between shareholder capital and insurance coverage in providing protection to creditors, see Radaszewski v. Telecom Corp., 981 F.2d 305 (8th. Cir. 1992). The court acknowledged the protection offered by insurance coverage to creditors, and proclaimed that "[i]nsurance is unquestionably relevant on the issue of 'undercapitalization.'" *Id.* at 310. What the court failed to appreciate was that insurance coverage is in fact much more effective in protecting creditors for the same amount of outlay by the corporation.

The nature of the inadequate capitalization inquiry is such that capital adequacy can only be determined in a rough and impressionistic manner.[208] This can be illustrated by a numerical example. If a corporation has outstanding liabilities of $1 million, and its equity capital is only $10,000, few courts would have trouble finding inadequate capitalization. However, if the initial capitalization had been $250,000, the courts would have much greater difficulty reaching the same conclusion, even though $250,000 is clearly insufficient to cover $1 million of liabilities. If the court were to find $250,000 inadequate, the inevitable question would be what about $260,000, or $300,000? Given that equity capital was never intended fully to cover corporate liabilities, an inquiry about capital adequacy probably cannot go much beyond identifying cases of egregious inadequate capitalization.[209]

In trying to identify egregious cases, the courts look for evidence of bad faith use of the corporate entity on the part of the shareholders.[210] In particular, the courts have placed great emphasis on evidence tending to show that the shareholders have deliberately exploited the corporate entity to the detriment of creditors.[211] In fact, it has been suggested that inadequate capitalization is but evidence of shareholder bad faith. In *Radaszewski*, the Eighth Circuit proclaimed that "the creation of an undercapitalized subsidiary justifies an inference that the parent is either deliberately or recklessly creating a business that will not be able to pay its bills or satisfy judgments against it."[212] Determination of bad faith, deliberateness, or recklessness, entails consideration of evidence of the subjective intent of the shareholders. If the inquiry is meant to be an impressionistic one, and the goal is only to identify cases of shareholder bad faith, there is little need for a clear quantitative rule. Judicial inability to articulate a clear threshold for capital adequacy is less perplexing and troubling than it may have initially seemed. Correctly understood, inadequate capitalization is not an objective criterion entailing a comparison between the corporation's initial capital and its outstanding liabilities. Instead, it is about ascertaining the good faith of the shareholders in their operation of the corporation, and in particular, in their decisions to incur liabilities.

---

[207] Leebron, *supra* note 38, at 1633-35. Professor Leebron in fact suggested such an obligation on corporations with respect to tort liabilities.

[208] Krendl & Krendl, *supra* note 11, at 37 (arguing that "courts are not totally competent to determine what adequate capitalization should be for a particular business").

[209] It is perhaps due to this inability to increase the rigor with which inadequate capitalization is analyzed that the courts have not attached as much weight to this factor as the commentators have believed. Thompson II, *supra* note 11, at 1065-67.

[210] For example, in *Automotriz del Golfo v. Resnick*, the court concluded that the corporation suffered from inadequate capitalization given an initial capital of $5,000 and a business of $100,000 to $150,000 in sales. 306 P.2d 1 (Cal. 1957). It may be argued that the court focused on the wrong variable. The comparison should not be made with sales, but with outstanding liabilities. However, the amount of sales probably has some correlation with outstanding liabilities.

Not only is this refashioned inadequate capitalization criterion more consistent with judicial practice in reality, it also comports with the formal and substantive nature of the corporate veil doctrine identified in Sections 0 and 0. As demonstrated in Section 0, the corporate veil doctrine is a standard-based exception to limited liability. A standard encapsulates fundamental values of a legal system such as fairness and good faith. A refashioning of the capital adequacy inquiry into a subjective one about the good faith of the shareholders is hence in line with the nature of the doctrine as a standard. It is also consistent with the doctrine's altruistic nature. Requiring shareholders to demonstrate good faith in their operation of the corporation reinforces the altruistic call for sharing and sacrifice and rejects the exclusive concern for one's self-interest that is the hallmark of individualism. Moreover, shareholders are unjustly enriched only if they have incurred liabilities with deliberate or reckless disregard for creditor interest. As argued earlier, one of the necessary consequences of limited liability is that shareholders generally will not bear responsibility for corporate liabilities beyond their equity investments in the firm. When outside creditors are left unpaid, shareholders are enriched, but not unjustly so. It is only when creditors are unpaid because liabilities were incurred with flagrant disregard for creditor interest that shareholders can be said to have unjustly enriched themselves.

---

[211] In *Costello v. Fazio*, an equitable subordination case, the court concluded that the corporation was undercapitalized when the initial capital was $6,000 and the current liabilities were $162,162. 256 F.2d 903, 905 (9th Cir. 1958). Moreover, the court attached considerable weight to the fact that the incorporators had withdrawn upon incorporation close to $40,000 of capital from the business, which previously had been operated as a partnership. *Id.* at 909 ("It is therefore certain, contrary to the finding just noted, that, in withdrawing this capital, Fazio and Ambrose did act for their own personal and private benefit.").

[212] Radaszewski v. Telecom Corp., 981 F.2d 305, 308 (8th Cir. 1992).

It remains to be set out how shareholder bad faith is to be determined. Even though a precise and quantitative comparison has been rejected as impractical, a determination of bad faith will still need to take into account the amount of initial capital and liabilities incurred by the corporation, which obviously are only known subsequently.[213] Moreover, if the question is whether the shareholders have demonstrated deliberate or reckless disregard for creditor interest when incurring a legal obligation, the temporal reference point should not be when the corporation fails, but when the obligation is incurred. The court should compare the corporation's outstanding liabilities with the initial capital when the legal obligation at issue is incurred. To sum up, capital adequacy should be determined by whether, in light of the initial capital, the shareholders incurred a legal obligation with the knowledge that the corporation was unlikely to be able to repay it, or with reckless disregard for the likelihood of repayment.

---

[213] For example, there is nothing inherently dishonest or reprehensible about a corporation with an initial capital of $25,000, if the corporation only incurs liabilities up to $50,000. The deliberate disregard for its ability to repay its debts can only be shown if it has taken on liabilities of a much larger amount. This obviously cannot be determined only with reference to the initial capitalization.

(iii) Different Analyses for Contract and Tort Cases

In determining capital adequacy, it is important to distinguish contract cases from tort cases. This is due to a fundamental difference between contractual and tort obligations. The former is incurred on a voluntary basis, whereas the latter is usually involuntary in character. This is of course true for the contractual creditor and the tort victim, but is in fact true for the corporation as well. A corporation voluntarily assumes a contractual obligation by entering into a contract with a lender, a supplier, or an employee. Contractual obligations are not forced upon a corporation against its will. The same cannot be said about tort obligation. Negligence-based tort obligations, which account for most corporate torts, are involuntarily incurred. In fact, they are usually assiduously avoided.

Given the control exercised by a corporation over the incurrence of contractual obligations, the shareholders have an *a priori* opportunity to evaluate its business operation and assess the likelihood that the corporation will be able to meet a particular contractual obligation. Therefore, a decision to incur a contractual obligation that grossly exceeds the corporation's capital base and ability to repay should be weighed heavily when assessing capital adequacy. The court may infer from this fact a deliberate disregard for the corporation's inability to repay its debts on the part of the shareholders. Their involuntary character means that the same inference cannot be hastily drawn about tort obligations.[214] What sort of inference can be drawn instead depends on the foreseeability of risk. If an activity is sufficiently high-risk, one may perhaps argue that the shareholders should have foreseen it. If a corporation is set up with a very small initial capital but is used to operate an ultra-hazardous business and the potential liability arising from it is enormous, then the shareholders can be said to have shown reckless disregard for the likelihood of repayment. In most other tort situations where the risk is relatively low, the same inference of recklessness may not be drawn and the mere incurrence of the tort obligation does not conclusively demonstrate inadequacy of capital.

---

[214] Some commentators have suggested that inadequate capitalization alone should suffice to justify veil piercing in tort cases. In light of the foregoing discussion, it would perhaps seem that this should apply in contract cases and not tort cases.

Even for contractual obligations, one should not be too quick to find inadequate capitalization from the mere incurrence of an obligation that grossly exceeds the corporation's capital base and ability to repay. It is important to recall that default risks are largely determined by the state of a corporation's business operations, not the size of its shareholder capital. Therefore, a contractual obligation may dwarf the corporation's capital base, but the shareholders may in good faith believe that the corporation will be able to repay it with the revenue generated from its operations. The corporation may well have been able to meet this obligation but for unforeseen circumstances in its operation that deprived it of substantial revenue. Inadequate capitalization thus should not be found if evidence shows that when the contractual obligation was incurred, there was at least some reasonable basis for the shareholders to believe that the corporation could generate sufficient revenue to cover the liability. This showing need not be high; some reasonable basis for belief would suffice. Otherwise the courts would have to venture into the unfamiliar territory of second-guessing business decisions *ex post*. In sum, a gross disparity between the initial capital of the corporation and its subsequent liabilities should only establish a presumption of inadequate capitalization. The court should then consider whether the shareholders had a reasonable basis to conclude that repayment was likely.

An adjustment needs to be made in order for this mode of analysis to apply to tort obligations. The need for adjustment again arises from the involuntary nature of tort obligations. In determining capital adequacy with respect to the incurrence of a contractual obligation, it makes sense for the courts to direct the inquiry to the time of incurrence of the obligation. This is because that is the time at which the shareholders make a conscious decision to take on the obligation. The situation is different for tort obligations. The shareholders do not make a decision to take on the obligation, but only to undertake an activity that may expose the firm to potential tort liabilities.

Therefore, when applying the foregoing mode of analysis to tort obligations, the courts should compare initial capital with outstanding liabilities at the time the corporation undertakes the activity that eventually gives rise to tort liability. That is the time at which the shareholders resolve to take on the potential exposure to tort liability.

(iv) Summing Up

Reformulation of the criterion of inadequate capitalization into an inquiry about good faith use of the corporate entity may seem surprising to some. Critics may argue that this modification entails a wholesale change in the nature of the criteria from one about the formal integrity of the corporation to one about the propriety of shareholder conduct. While this reformulation clearly shows that the criterion belongs to the impropriety prong rather than the formalities prong, it does not change the nature of the inquiry. Unlike other criteria under the formalities prong such as observance of corporate formalities or overlap in corporate personnel, inadequate capitalization was never truly about the form of the corporation. It was always about economic substance. A proper classification should thus place it under the impropriety prong.

This classification is consistent with the reformulation of the doctrine under a shareholder unjust enrichment rationale. The result is that most of the factors under the formalities prong have been de-emphasized and the main inquiry of the doctrine is shifted to the identification of shareholder unjust enrichment under the impropriety prong. As inadequate capitalization is clearly an instance of enrichment, it rightfully should be included under the impropriety prong. Substantively, an inadequate capitalization inquiry that focuses on the extent to which the shareholders deliberately or recklessly incur legal obligations to the detriment of creditor interest brings it much closer in line with the rationale of the corporate veil doctrine. Unjust enrichment results if the shareholders have incurred a legal obligation with deliberate or reckless disregard for the creditors' interests.

To recap the discussion in this section, inadequate capitalization should be considered under the corporate veil doctrine, and in particular under the impropriety prong, in light of the doctrine's shareholder unjust enrichment rationale. Efforts to formulate a quantitative standard for determining capital adequacy is likely to be misguided given that the courts are only likely to intervene in egregious cases and are ultimately more concerned about whether the shareholders have exploited the corporate entity in good faith. Instead, the inquiry should be refashioned explicitly to focus on the shareholders' good faith in their operation of the corporation. Lastly, contract and tort cases should be treated differently in the analysis. The need for different treatment is largely due to the voluntary nature of contractual obligations and involuntary nature of tort ones.

2. Other Improper Conduct

Given the variety of improper conduct, it is probably impossible to analyze each of them in great detail. The main conclusion from this article is that when deciding what sort of conduct justifies veil piercing, the courts should focus on whether the conduct results in shareholder unjust enrichment at the expense of the corporate creditors. This rationale for the corporate veil doctrine should help to sharpen the analysis and provide greater predictability to future litigants.

This can be illustrated by two examples, asset stripping and deliberate unprofitable operation of a corporation. Similar to inadequate capitalization, shareholder misappropriation of corporate assets clearly results in shareholder unjust enrichment and justifies veil piercing. For example, in *Henderson v. Rounds & Porter Lumber Co.*, the court pierced the veil on the grounds that the corporate parent had pilfered the subsidiary's assets.[215] The court concluded that such asset stripping was tantamount to inadequate capitalization. Although the court did not spell it out explicitly, the analogous nature of the two is due to the fact that they both result in unjust enrichment of shareholders. Operation of a corporation in a deliberately unprofitable manner similarly results in shareholder unjust enrichment.[216] One example of such operation is when a parent corporation sets up a subsidiary with the sole purpose of supplying a raw material or an input of production at cost or a loss-making price.[217] The profit that the subsidiary could have made if its output was sold at the prevailing market price would have been available to satisfy any outstanding claims it may have against it. Instead, the parent requires the subsidiary to sell the output at cost or even below cost, which means that the subsidiary will never make a profit, thereby unjustly enriching the parent. In unjust enrichment terms, there is little practical difference between this mode of operation, asset stripping, and inadequate capitalization. All three should justify veil piercing.

---

[215] 99 F. Supp. 376 (W.D. Ark. 1951).

C. Remedy

One final issue concerning recasting the corporate veil doctrine in terms of shareholder unjust enrichment is the measure of recovery. In a standard unjust enrichment claim, the measure of recovery is the amount of the defendant's gain from the unjust enrichment.[218] Transposing it to the veil piercing context, if the improper conduct involved is inadequate capitalization or asset stripping, the measure of recovery should be the shortfall in capitalization or the amount of the misappropriated assets. This would create obstacles to the doctrine as an exception to limited liability, under which the standard remedy is the corporation's unpaid liability. This apparent contradiction between shareholder liability under the corporate veil doctrine and its unjust enrichment rationale can be reconciled. While the standard measure of recovery under unjust enrichment is the defendant's gain, there are circumstances under which the measure of recovery becomes the plaintiff's loss due to difficulty in quantifying the value of the defendant's gain or the fact that the defendant's conduct has destroyed value.[219]

---

[216] Landers, *supra* note 68, at 621 (noting that deliberate unprofitable operation of subsidiary justifies veil piercing against parent).

[217] *See, e.g.*, United States v. Reserve Mining Co., 380 F. Supp. 11 (D. Minn. 1974).

[218] Laycock, *supra* note 114, at 1279.

The defendant's gain may be difficult to quantify in some veil piercing situations. In inadequate capitalization cases, the defendant' gain is the shortfall in capitalization. However, the court's inability to come up with more precise measures of capital adequacy means that they are unable to determine the amount of capital shortfall with confidence. The refashioning of the criterion into an inquiry about good faith operation of the corporation instead of a quantitative comparison between equity capital and outstanding liabilities also means that capital shortfall is no longer an appropriate measure of recovery. In cases involving commingling of shareholder and corporate assets, the finances of the culpable shareholder and the corporation may become so hopelessly intertwined that the court is unable to determine the precise amount by which the shareholder has benefited from corporate assets. This again means that the court will not be able to quantify the defendant's gain. Finally, in deliberate unprofitable operation cases, it may be impossible for a variety of reasons to determine the market price for the under-priced input of production, or the subsidiary may have been exploited in other ways that are difficult to quantify. This will again prevent the court from quantifying the defendant's gain. Given the complex nature of the facts in many corporate veil cases, especially the extent to which the corporation's and the culpable shareholder's affairs and finances are intertwined, this difficulty in quantification is likely to be common. This difficulty is compounded by the fact a corporate veil case may involve not one instance but a range of improper conduct by the shareholders. Calculating the defendant's gain from each of them may be exceedingly complex and time-consuming. Measuring recovery by the plaintiff's loss will sidestep these myriad problems.

In addition, there are reasons to believe that in many corporate veil cases, shareholder misconduct destroys value in the corporation and results in a loss to the plaintiff that is greater than the defendant's gain from its misconduct. For example, asset stripping harms the corporation not only through the loss of the pilfered assets, but also in other less direct ways. The impact of misappropriation on the corporation's ability to repay its creditors is likely to be greater than the amount of lost assets. A corporation known to have been a victim of shareholder misconduct is likely to lose its creditworthiness and may become subject to loan recall by its financial creditors. Suppliers may also become unwilling to supply goods or services on credit. Employees may start to leave the firm. Even if the misconduct is not known to the creditors, pillage by the shareholders may leave the firm in such poor financial health that it is unable to pursue otherwise profitable business opportunities. These opportunities would have brought in revenue that could have been used to pay off the creditors. The same scenario may arise under inadequate capitalization. Here, even though the shareholders did not pilfer the firm's capital, they endowed the firm with such a small starting capital that its ability to borrow or to pursue profitable business opportunities is greatly impeded. In both instances, the improper conduct that results in unjust enrichment inflicts greater loss on the plaintiff than it benefits the defendant. Measuring recovery by the defendant's gain would under-compensate the plaintiff.

---

[219] *Id.* at 1285, 1287 ("In some cases, unjust enrichment best explains substantive liability, but plaintiff's loss is the most appropriate measure of recovery.").

The foregoing discussion hopefully has demonstrated that measuring recovery by the defendant's gain is likely to be unsatisfactory in most corporate veil cases. One may argue that the measure of recovery can be adjusted on a case-by-case basis depending on the difficulty of quantification of the defendant's gain in a particular case. This, however, will introduce needless complication to what is already a highly confusing area of law. There are bound to be cases in which the defendant's gain exceeds the plaintiff's loss and vice versa. If a uniform approach must be adopted, the choice is between overcompensating the plaintiff in some cases and under-penalizing the defendant in other cases. Given the clear culpability of the defendant in most corporate veil cases, there are compelling policy reasons to choose the former. Therefore, even though the standard measure of recovery in unjust enrichment cases is the defendant's gain, the appropriate remedy in corporate veil cases should be the plaintiff's loss, which is the unmet liability. Imposing shareholder liability in corporate veil cases is hence entirely consistent with its unjust enrichment rationale.

CONCLUSION

The approach taken in this article to the corporate veil doctrine is distinctly more theoretical than is common in the existing literature. Applying the framework from Professor Kennedy's seminal article on the form and substance of private law adjudication, it established the formal and substantive relationships between the corporate veil doctrine and limited liability and confirmed that the substantive basis of the doctrine is one of altruism. Building on this altruistic foundation, the article demonstrated that the principal objective of the corporate veil doctrine is the prevention of shareholder unjust enrichment at the expense of the corporation's creditors. It further argued that much of the current confusion in the doctrine can be dispelled if it is reoriented to focus on shareholder unjust enrichment. Tangential considerations under the formalities prong of the doctrine such as observance of corporate formalities, shareholder domination, overlap in corporate personnel, and substantive separation have largely obscured the analysis and should be de-emphasized. Instead, the inquiry should focus on the existence of shareholder unjust enrichment under the impropriety prong. This will bring the doctrine closer to its proper theoretical foundation and help the courts to identify what truly drives veil piercing decisions.

AGAINST THE CREATION MYTH OF TEXTUALISM: THEORIES OF CONSTITUTIONAL INTERPRETATION IN THE NINETEENTH CENTURY

John P. Figura[*]

INTRODUCTION 588
I. THE CONTEMPORARY DEBATE: PURPOSIVISM VERSUS TEXTUALISM 591
II. JOSEPH STORY AND PLAIN-MEANING PURPOSIVISM 596
  A. Origins: Blackstone and Marshall 597
  B. Joseph Story's Plain-Meaning Purposivism 600
  C. Following Story's Lead: Plain-Meaning Purposivism to 1868 605
    1. The Surprisingly Aggressive Purposivism of Theodore Sedgwick 606
    2. Thomas Cooley: The Progressive Version 608
  D. The Last of the Story Disciples? 611
  E. A Not-Quite-So-Plain Meaning School 612
III. A MORE DIRECT PURPOSIVISM: THE RETREAT OF THE PLAIN MEANING RULE 613
  A. Early Conventional Constitutional Purposivism 614
  B. The Next Wave: The Plain Meaning Rule Rescinded 617
IV. EVOLUTIONARY CONSTITUTIONALISM 621
  A. Sources 622
    1. The Evolutionary Principle 622

---

[*] Law Clerk to the Honorable Jack B. Weinstein, United States District Court for the Eastern District of New York. J.D., with honors, Order of the Coif, Emory Law School. M.A., English, The Ohio State University. B.A., Augustana College. The author wishes to thank Victoria Nourse, Visiting Professor of Law, Georgetown University, Burrus-Bascom Professor of Law, University of Wisconsin, for posing the question that launched this project and for her invaluable guidance throughout its execution. Thanks are also due to Professors H. Jefferson Powell, Robert Schapiro, Charles Shanor, Stephen Siegel, and Adam Winkler for their thoughtful comments and words of encouragement.

    2. The Anti-Constitutionalists 623
    3. The Lieber Revival 626
  B. Social Darwinist Constitutionalism: The Conservatives 627
    1. William Gardiner Hammond 627
    2. Francis Wharton 629
    3. Christopher Tiedeman 631
  C. Reform Darwinist Constitutionalism: The Progressives 634
CONCLUSION 636

INTRODUCTION

Historical precedent plays a significant role in the ongoing debate between textualists and purposivists. As the familiar historical account has it, purposivism is a twentieth-century phenomenon, a digression from a more deep-seated American tradition of text-focused interpretation. Seen in this light, purposivism is but a misguided experiment and textualism—especially as employed in originalist constitutional interpretation—a sort of renaissance, a return to the wiser, more sensible base line of text-based American legal interpretation.[1]

If this account were true, the nineteenth century would reveal a strong tradition of text-based constitutional interpretation. But a close look at the nineteenth century's interpretive theories, as put forth in legal treatises of the era, reveals nothing of the sort. The century was dominated by purpose-based, not text-based, approaches to the Constitution. In this article, I identify three dominant interpretive tendencies: the plain-meaning purposivism of mid-century, as defined by Joseph Story; the conventional purposivism that arose after the Civil War; and the late-century, anti-originalist school of evolutionary constitutionalism. None of these was more than superficially textualist, and none was consistent with the historically conservative brand of originalism espoused by textualists such as Justice Antonin Scalia.

---

[1] *Infra* text accompanying notes 28–36.

This history indicates not only that the nineteenth century was a mostly purposivist era but that purposivism was, in the nineteenth century, the textually conservative choice, promising greater fidelity to the text of the Constitution and the understandings of the founding generation than any extant alternative. These lessons are particularly relevant today, as courts and commentators move toward a broad acceptance of textualism and a rejection of purposivism.[2]

By focusing on constitutional treatises, I have sought to isolate the interpretive rules that nineteenth-century courts followed. Of course, it is case law that matters most, not scholarly works. But nineteenth-century treatises were primarily practical works; their authors largely eschewed abstract theory and attempted to describe law as it was actually applied.[3] Their work is therefore a useful—if imprecise—mirror of judicial practice.

Nineteenth-century treatise writers' rules for constitutional interpretation are to a great extent applicable to statutory interpretation as well. Nineteenth-century thinkers often wrote of the two types of interpretation not as separate spheres but as following the same set of rules.[4] This should not be surprising; the first authoritative text used by American jurists for interpretation of any sort was Blackstone's *Commentaries*, which was written for a country with no written constitution. Over time, Americans would develop separate schemes for statutory and constitutional interpretation.[5] To avoid confusion, I have indicated in the text or notes those passages where I have inferred a rule of constitutional interpretation from a rule of general or statutory interpretation.

---

[2] *See infra* text accompanying note 36.

[3] *See* LAWRENCE M. FRIEDMAN, A HISTORY OF AMERICAN LAW 244–45 (3d ed. 2005) (stating that nineteenth-century American legal treatise writers eschewed prescriptive, academic approaches, instead attempting to instruct working lawyers on how law was actually practiced). *See also* FRANCIS R. AUMANN, THE CHANGING AMERICAN LEGAL SYSTEM 127-28 (Da Cappo Press 1969) (1940) (stating that the widespread distribution of high-quality treatises by authors such as Story and James Kent helped unify American legal doctrines); John W. Head, *Codes, Cultures, Chaos, and Champions: Common Features of Legal Codification Experiences in China, Europe, and North America*, 13 DUKE J. COMP. & INT'L L. 1, 72-73 (2003) (same).

[4] *See, e.g., infra* notes 117, 121, 168, and 169.

[5] *See, e.g., infra* text accompanying notes 92–93 (discussing Henry Sedgwick's divergent views of statutory and constitutional interpretation).

In Part I, I address the writings of Joseph Story and the other adherents of what I call plain-meaning purposivism—the group that has been described as the "plain meaning school." Because of its use of the plain meaning rule, which states that a court should consult extratextual sources only when the plain meaning of a provision is unclear, this approach is often seen as a precursor to contemporary textualism. But the nationalist Story and his mid-century followers subordinated their seemingly text-based approach to a set of substantive—and only loosely originalist—presumptions about the role of the federal government. Because these presumptions were based on extratextual deductions of constitutional principles, this school took a fundamentally purposivist approach to constitutional interpretation.

Part II addresses a more conventionally purposivist approach that flourished from the late 1860s through the end of the century. Adherents of this approach softened or eliminated the plain meaning rule that had figured so prominently in Joseph Story's approach. Instead, they advised courts to freely consult extratextual materials, both of historical and contemporary origin, even in the face of clear constitutional language. Like purposivists today, these writers were not a united school of thought but an ideologically diverse group united by little more than their orientation to the text.

In Part III, I discuss evolutionary constitutionalism. Melding the theories of Edmund Burke and Herbert Spencer, these theorists abandoned originalism and embraced not the text but the notion of an unwritten, evolving Constitution. In their view, courts should settle constitutional questions not by original intent or the meaning of the text but by the character of the living nation and the needs of its people. This was primarily a conservative, pro-industrial movement, but its evolutionary, non-purposivist model would be adapted to progressive ends after the turn of the century.

I. THE CONTEMPORARY DEBATE: PURPOSIVISM VERSUS TEXTUALISM

This article looks at nineteenth-century theories of constitutional interpretation primarily through the lenses of textualism and purposivism. Purposivists believe judges should interpret ambiguous laws—usually statutory provisions—with an eye toward serving the purposes for which they were enacted.[6] In determining purpose, purposivists look to numerous sources: statutory context, legislative history, the mischief that was to be remedied by the provision, and the text itself.[7]

Textualism—or "the new textualism," as Professor William Eskridge called it in 1990—emerged in the 1980s as a reaction against the perceived excess of purposivism. In the view of textualists, purposivism allowed interpreters—particularly the members of the much-maligned Warren Court—to substitute their own policy preference for the meanings of the text.[8] They charged that this undermined the rule-like nature of law[9] and unconstitutionally expanded the power of unelected and unaccountable judges.[10] Textualists view intent itself as inscrutable at best[11] and inherently nonexistent at worst.[12] Even if intent could be ascertained, to treat it as law—in place of the text of a provision—would violate the Constitution.[13]

---

[6] The fathers of twentieth-century purposivism, Henry Hart and Albert Sacks, stated that a statutory interpreter must:

> Decide what purpose ought to be attributed to the statute and to any subordinate provision of it which may be involved; and then [i]nterpret the words of the statute immediately in question so as to carry out the purpose as best it can, making sure, however, that it does not give the words either—(a) a meaning they will not bear, or (b) a meaning which would violate any established policy of clear statement.

HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW 1374 (William N. Eskridge, Jr. & Philip P. Frickey, eds., 1994) (internal line breaks omitted).

[7] *Id.* at 1379.

[8] William N. Eskridge, Jr. & Philip P. Frickey, *Introduction* to HART & SACKS, *supra* note 6, at cv.

[9] Frank H. Easterbrook, *The Role of Original Intent in Statutory Construction*, 11 HARV. J. L. & PUB. POL'Y 59, 65 (1988) [hereinafter Easterbrook, *Original Intent*].

[10] Antonin Scalia, *Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws*, *in* A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 17-18 (1997).

[11] *See* Easterbrook, *Original Intent*, supra note 9, at 60 (stating that purposivism requires judges "to figure out the mental pattern of the legislature"); *id.* at 65 ("Words

Instead of intent, textualists focus on the text itself.[14] They apply what Professor Eskridge has called the "hard" plain meaning rule, which states that there is no need to look beyond the text when the textual meaning is clear.[15] Purposivists, by contrast, apply a "soft" plain meaning rule, which allows interpreters to consult extratextual sources freely and to use them in contradiction of clear statutory language.[16] While purposivists concern themselves with the policy context of ambiguous provisions, textualists limit their inquiry to what Professor John Manning has called semantic context—that is, the generally understood meanings of the words on the page.[17] Therefore, instead of consulting legislative history,[18] they rely on dic-

---

appeal to the reasonable man of tort law; private language and subjective intents should be put aside."); Frank H. Easterbrook, *Statutes' Domains*, 50 U. CHI. L. REV. 533, 551 (1983) ("Few of the best-intentioned, most humble, and most restrained among us have the skills necessary to learn the temper of times before our births, to assume the identity of people we have never met, and to know how 535 disparate characters from regions of great political and economic diversity would have answered questions that never occurred to them.").

[12] *See* Frederick Schauer, *Statutory Construction and the Coordinating Function of Plain Meaning*, 1990 SUP. CT. REV. 231, 232 (1990) (The search for actual legislative intent is "a theoretically optimizing but likely self-defeating search for first-best solutions by multiple decisionmakers with different goals and different perspectives"); Easterbrook, *Statutes' Domains*, *supra* note 11, at 547 ("Because legislatures comprise many members, they do not have 'intents' or 'designs,' hidden yet discoverable.").

[13] *See* Easterbrook, *Original Intent*, *supra* note 9, at 60.

[14] *See* William N. Eskridge, Jr., *Textualism, the Unknown Ideal?*, 96 MICH. L. REV. 1509, 1511 (1997) (reviewing ANTONIN SCALIA, A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW (1997)).

[15] *See* William N. Eskridge, Jr., *The New Textualism*, 37 UCLA L. Rev. 621, 656-57 (1989) [hereinafter Eskridge, *New Textualism*]; John F. Manning, *Textualism and the Equity of the Statute*, 101 COLUM. L. REV. 1, 7 (2001) [hereinafter Manning, *Equity of the Statute*] ("[T]extualists believe that when a statutory text is clear, that is the end of the matter").

[16] *See* Eskridge, *New Textualism*, *supra* note 15, at 626 (citing Church of the Holy Trinity v. United States, 143 U.S. 457 (1892)) (stating that *Holy Trinity* indicates the soft plain meaning rule).

[17] John F. Manning, *What Divides Textualists from Purposivists?*, 106 COLUM. L. REV. 70, 92 (2006) [hereinafter Manning, *What Divides Textualists*].

[18] Eskridge, *New Textualism*, *supra* note 15, at 623 ("Legislative history should not even be consulted to confirm the apparent meaning of a statutory text.").

tionaries, other statutes, and common law precedent.[19] The result, in the view of textualists, is less judicial discretion, more predictable results, and more fidelity to the work of democratically accountable representatives.[20]

A particular problem arises when seemingly textualist interpretation—adherence to the text instead of extratextual indicia of meaning—is used to determine lawmakers' intent. Textualism, at least as we know it today, is philosophically inconsistent with the search for intent. If an interpreter relies on text as a means to achieve intentionalist ends, the textual focus is a matter not of principle but of expediency. In many cases, it is a response to the perceived unavailability of reliable extratextual indicia of lawmakers' intent. Such an interpretive approach is probably best thought of not as textualism but as intentionalist pseudo-textualism.

Both textualism and purposivism apply to constitutional interpretation as well as statutory interpretation. Purposivists believe courts should interpret the Constitution to serve the document's overarching purposes, structure, and standards.[21] So, instead of limiting their inquiry to the text, they look to abstract principles—say, democracy, liberty, equality, or the protection of private property.[22] Textualists also see their approach as applicable to constitutional questions.[23] They generally shun the notion of determining the intent of the framers and instead look to sources, such as dictionaries or *The Federalist* papers, that would indicate the public meaning of the words in the Constitution.[24]

The textualist/purposivist divide only partially overlaps with a more prominent debate in constitutional interpretation: that between originalists and non-originalists. Originalism has in the last decade proven to be a big tent, encompassing everything from the attempt to recapture the content of the framers' opinions to the commitment to honor the framers' spirit of philosophical inquiry.[25] For the purposes of this discussion, it is important to maintain the distinction between textualist originalism, as typified by the constitutional interpretation of Justice Scalia,[26] and conceptual or abstract originalism, which I would suggest is generally purposivist.[27] I will return to this distinction later in the article as I discuss the three interpretive approaches that predominated in the nineteenth century.

---

[19] Manning, *What Divides Textualists*, *supra* note 17, at 92.

[20] Scalia, *supra* note 10, at 16-18.

[21] Vicki C. Jackson, *Multi-Valenced Constitutional Interpretation and Constitutional Comparisons: An Essay in Honor of Mark Tushnet*, 26 QLR 599, 611 (2007).

[22] *Id.* at 611–12.

[23] Scalia, *supra* note 10, at 37 ("The [constitutional] problem is distinctive," Justice Scalia wrote, "not because special principles of interpretation apply, but because the usual principles are being applied to an unusual text.").

[24] *Id.* at 38.

Accompanying the theoretical argument between textualism and purposivism is a historical narrative that figures textualism as a renaissance, a new-and-improved return to an older, wiser approach to interpretation that proliferated before the twentieth century. In the view of Professor Leonard Levy, what we know today as textualism dates to the Founders.[28] Judge Frank Easterbrook has labeled the father of American jurisprudence, John Marshall, "the first great textualist."[29] And numerous commentators have linked modern textualism to the nineteenth-century "plain meaning" school, either describing it as a more sophisticated version of that school[30] or asserting that the two are one and the same.[31]

---

[25] SOTIRIOS A. BARBER & JAMES E. FLEMING, CONSTITUTIONAL INTERPRETATION: THE BASIC QUESTIONS 79, 83–84 (2007). *See also* Scalia, *supra* note 10, at 38.

[26] *See, e.g.*, Scalia, *supra* note 10, at 37.

[27] *See* BARBER & FLEMING, *supra* note 25, at 83–84.

[28] LEONARD W. LEVY, ORIGINAL INTENT AND THE FRAMERS' CONSTITUTION 2 (1988) (stating that the Framers, who did little to memorialize their proceedings, wished future interpreters to rely not on their intentions but on "the text of the Constitution, construed in the light of conventional rules of interpretation, the ratification debates, and other contemporary expositions").

[29] Frank H. Easterbrook, *Textualism and the Dead Hand*, 66 GEO. WASH. L. REV. 1119, 1122 (1997) [hereinafter Easterbrook, *Dead Hand*].

[30] *E.g.*, Eskridge, *New Textualism*, *supra* note 15, at 623 n.11 ("What is 'new' about the new textualism is its intellectual inspiration: public choice theory, strict separation of powers, and ideological conservatism.").

[31] Martin H. Redish & Dennis Murashko, *The Rules Enabling Act and the Procedural-Substantive Tension: A Lesson in Statutory Interpretation*, 93 MINN. L. REV. 26, 49 n.99 (2008) (criticizing the "new textualism" label for suggesting that textualism is a recent development).

By contrast, textualists have popularized a conception of purposivism as relatively young, originating around the time of the (infamous) 1892 decision of *United States v. Holy Trinity Church*.[32] As Justice Scalia figures it, purposivism as we know it today achieved prevalence during the Hart-and-Sacks era of the 1940s and 1950s, when judges began making extensive (and, in the view of textualists, misguided) use of legislative history, which had not been consistently available previously.[33]

Justice Scalia called in 1995 for an end to the "brief and failed experiment" with legislative history.[34] Today, it appears that his wish has been fulfilled to a great extent, as commentators appear ready to announce the death of purposivism. In the words of Professor Jonathan Molot, "textualism has been so successful in altering the views of even nonadherents"[35] that it is now safe to say that "we have all become textualists."[36]

While no one suggests that our current interpretive choices should be determined by those of nineteenth-century jurists, this historical narrative has a certain power. If textualism is the de facto baseline of American constitutional and statutory interpretation—and if purposivism has been a mere brief but failed experiment—then embracing textualism and abandoning purposivism seems a careful, conservative option.[37]

This narrative may be powerful, but it is not accurate, as this article will show. Most nineteenth-century treatise writers who addressed constitutional interpretation—and particularly those whose guidelines resemble today's textualism—espoused purpose-based interpretation. Their conception of meaning was generally broad; instead of limiting themselves to a narrowly originalist perspective—where they would be bound by what the original drafters of an instrument would have said if faced with a precise contemporary issue—they instructed judges to reason from principle and purpose. In a word, they were purposivists.[38]

---

[32] *E.g.*, Scalia, *supra* note 10, at 18–21 (citing Church of the Holy Trinity v. United States, 143 U.S. 457 (1892)); Manning, *Equity of the Statute*, *supra* note 15, at 13-14 (naming *Holy Trinity* as the leading case in purposivist interpretation). *See also generally* Adrian Vermeule, *Legislative History and the Limits of Judicial Competence: The Untold Story of* Holy Trinity Church, 50 STAN. L. REV. 1833 (1997) (discussing *Holy Trinity* as demonstrating the errors of purposivism).

[33] Scalia, *supra* note 10, at 30.

[34] *Id.* at 36.

[35] Jonathan T. Molot, *The Rise and Fall of Textualism*, 106 COLUM. L. REV. 1, 30 (2006).

[36] *Id.* at 36. *See also* Jonathan R. Siegel, *Textualism and Contextualism in Administrative Law*, 78 B.U. L. REV. 1023, 1057 (1998) ("In a significant sense, we are all textualists now.").

[37] *But see generally* Philip P. Frickey, *Interpretive-Regime Change*, 38 LOY. L.A. L. REV. 1971, 1971 (2004) (discussing the policy costs of embracing textualism after a fifty-year experience with modern purposivism).

II. JOSEPH STORY AND PLAIN-MEANING PURPOSIVISM

> [A] constitution of government does not, and cannot, from its nature, depend in any great degree upon mere verbal criticism, or upon the import of single words. . . . [W]e should never forget, that it is an instrument of government we are to construe; and . . . that must be the truest exposition, which best harmonizes with its design, its objects, and its general structure.
>
> —*Joseph Story*[39]

Commentators have discussed the nineteenth-century plain-meaning interpreters as formalists, woodenly applying the law in an effort at detached, scientific objectivity.[40] Professor Eskridge has described this group as a predecessor to contemporary textualism, but with a looser plain meaning rule and without the intellectual motivations of today's textualists.[41] Some commentators have disputed this distinction and labeled them textualist.[42]

---

[38] *See* Martin H. Redish & Theodore T. Chung, *Democratic Theory and the Legislative Process: Mourning the Death of Originalism in Statutory Interpretation*, 68 TUL. L. REV. 803, 813, 815 (1994); ("Intentionalism asks how the enacting legislature would have decided the interpretive question facing the court. . . . [P]urposivism calls on judges to identify the statute's broader purposes and to resolve the interpretive question in light of those purposes.").

[39] JOSEPH STORY, 1 COMMENTARIES ON THE CONSTITUTION OF THE UNITED STATES 441 (Boston, Hilliard, Gray, & Co. 1833).

[40] *E.g.*, Seth Barrett Tillman & Steven G. Calabresi, Debate, *The Great Divorce: the Current Understanding of the Separation of Powers and the Original Meaning of the Incompatibility Clause*, 157 U. PA. L. REV. PENNUMBRA 134, 150 (2008), http://www.pennumbra.com/debates/pdfs/GreatDivorce.pdf (characterization by Tillman of Joseph Story's interpretive method as a "very 'wooden' (i.e., Anglo-American, traditional, eighteenth-century, literalistic) textualism"); H. Jefferson Powell, *Joseph Story's Commentaries on the Constitution: A Belated Review*, 94 YALE L.J. 1285, 1296–97 (1984) (stating that Story took a scientific view of the Constitution and perceived government as a perfectable machine).

But none of these descriptions properly applies to plain-meaning constitutional interpretive theory. Adherents of this method approached the Constitution with neither a search for objective truth nor a desire to uphold the text at all costs. They were guided by policy concerns, particularly their view that the Founders intended to create a potent central government. Because they turned to the text only with this understanding of principles in mind, the plain-meaning constitutional interpreters of the nineteenth century were fundamentally purposivist.

A. Origins: Blackstone and Marshall

In his *Commentaries on the Laws of England*, first published in 1765, Sir Henry Blackstone memorialized the two most durable rules in legal interpretation: the intention rule and the plain meaning rule. But he made clear that it was intention, not a textual focus, that was to control, when he offered the following as an overall maxim: "The fairest and most rational method to interpret the will of the legislator, is by *exploring his intentions* at the time when the law was made, by signs most natural and probable."[43] The plain meaning rule appears as the first of a series of more specific guidelines, subordinated to that overriding mission: "Words are generally to be understood in their usual and most known signification . . . ."[44] Statutory language, then, is a mere vehicle for legislative intent.

---

[41] Eskridge, *New Textualism*, *supra* note 15, at 623 n.11.

[42] *E.g.*, Redish & Murashko, *supra* note 31, at 49 n.99 (criticizing the "new textualism" label for suggesting that textualism is a recent development).

[43] 1 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 59 (Oxford, Clarendon Press 1765) (emphasis altered from original).

[44] *Id.* The second rule in the series is a rule on extratextual sources. *See id.* at 60 ("If words happen to be still dubious, we may establish their meaning from the context"); *see also id.* (stating that context includes other words within the document as well as "other laws, that are made by the same legislator, that have some affinity with the subject, or that expressly relate to the same point").

Blackstone followed this with a series of purposivist guidelines. He wrote that a law is "always" to be understood with regard to its "subject matter"—that is, the mischief the law was meant to address.[45] More broadly, readers should interpret dubious words by considering "the reason and spirit" of a law; "when this reason ceases," he wrote, "the law itself ought likewise to cease with it."[46]

Blackstone's treatise was an immense success in the United States and played a major role in the development of early American jurisprudence.[47] Its American readership grew when the scholar St. George Tucker, a Jeffersonian Republican,[48] published a five-volume edition adapted to the laws of the United States and Virginia in 1803.[49] Tucker's adaptation copied Blackstone's rules of statutory interpretation verbatim.[50]

It would be decades before the United States saw an authoritative general legal treatise of its own,[51] but we can turn to the opinions of Chief Justice John Marshall for an indication of how Blackstone's formula was applied in constitutional interpretation.[52] Numerous scholars have characterized Marshall as either an early textualist or a precursor to textualism,[53] but his textual focus was offset by a strong, Blackstonian intentionalism that is at odds with contemporary textualism. He made this clear in *McCullough v. Maryland*, where in explaining his broad conception of the Necessary and Proper Clause, he reminded his readers, "[W]e must never forget that it is a *constitution* we are expounding."[54] That case involved nothing less than the:

---

[45] *Id.*

[46] *Id.* at 61. Blackstone's fourth rule concerned the absurdity doctrine. *Id.* at 60.

[47] FRIEDMAN, *supra* note 3, at 69.

[48] Mark Douglas McGarvie, *Transforming Society Through Law: St. George Tucker, Women's Property Rights, and an Active Republican Judiciary*, 47 WM. & MARY L. REV. 1393, 1397 (2005).

[49] Davison M. Douglas, *Foreword: The Legacy of St. George Tucker*, 47 WM. & MARY L. REV. 1111, 1112-13 (2006).

[50] *Compare* ST. GEORGE TUCKER, 1 BLACKSTONE'S COMMENTARIES: WITH NOTES OF REFERENCE, TO THE CONSTITUTION AND LAWS, OF THE FEDERAL GOVERNMENT OF THE UNITED STATES; AND OF THE COMMONWEALTH OF VIRGINIA 59-60 (Philadelphia, William Young Birch & Abraham Small 1803), *with* BLACKSTONE, *supra* note 43, at 59–60.

[51] Before Joseph Story's *Commentaries*, the most significant legal work of the early nineteenth century was James Kent's general treatise. Carl F. Stychin, *The Commentaries of Chancellor James Kent and the Development of an American Common Law*, 37 AM. J. LEGAL HIST. 440, 443 (1993) (describing the widespread distribution of Kent's treatise). It, however, provided no systematic rules for constitutional interpretation. *See generally* JAMES KENT, 1 COMMENTARIES ON AMERICAN LAW (New York, O. Halsted 1826). Story's work eclipsed Kent's as the authoritative treatment of constitutional law. John H. Langbein, *Chancellor Kent and the History of Legal Literature*, 93 COLUM. L. REV. 547, 592 (1993).

[52] Blackstone did not suggest applying formal rules of textual interpretation to a constitution; he designed his rules for statutory interpretation. *See generally* BLACKSTONE, *supra* note 43.

> [E]xecution of those great powers on which the welfare of a nation essentially depends. *It must have been the intention of those who gave these powers,* to insure, so far as human prudence could insure, their beneficial execution. This could not be done, by confiding the choice of means to such narrow limits as not to leave it in the power of congress to adopt any which might be appropriate, and which were conducive to the end.[55]

This intentionalist view calls into question the seemingly textualist dicta in *Sturges v. Crowninshield*. In that decision, Marshall wrote that:

> [A]lthough the spirit of an instrument, especially of a constitution, is to be respected not less than its letter, yet the spirit is to be collected chiefly from its words. It would be dangerous

---

[53] *E.g.*, Steven G. Calabresi, *The Tradition of the Written Constitution: Text, Precedent, and Burke,* 57 ALA. L. REV. 635, 638-39 (2005) (stating that Marshall's opinion in *Marbury v. Madison* pointed the way toward contemporary textualism); John F. Manning, *The Absurdity Doctrine*, 116 HARV. L. REV. 2387, 2388 (2002) (stating that the Marshall Court "entrenched the primacy of the text in matters of statutory interpretation"); Manning, *Equity of the Statute*, *supra* note 15, at 100-02 (discussing Marshall's establishment of the faithful agent theory as a hallmark of American statutory interpretation); Easterbrook, *Dead Hand*, *supra* note 29, at 1122-23 (applauding Marshall for relying "squarely on constitutional text—not on imputed intent, not on *The Federalist*, not on the debates in the Convention (which had been kept secret), not on the debates of the ratifying conventions . . . and not even on the opinions he had written for the Supreme Court") (footnotes omitted).

[54] McCulloch v. Maryland, 17 U.S. 316, 407 (1819).

[55] *Id.* at 415 (emphasis added). *See also* John Marshall, *A Friend of the Constitution*, ALEXANDRIA GAZETTE, July 2, 1819, *reprinted in* JOHN MARSHALL'S DEFENSE OF *MCCULLOUGH V. MARYLAND* 155, 167 (Gerald Gunther ed. 1969) (asserting that the common law provides "the most complete evidence that the *intention* is the most sacred rule of interpretation").

> in the extreme, to infer from extrinsic circumstances, that a case for which the words of an instrument expressly provide, shall be exempted from its operation.[56]

In light of *McCulloch*, it would perhaps be wise to read the "spirit" in *Sturges* as a synonym for "intention"—in which case Marshall's seeming textualism, like Blackstone's plain meaning rule, was subordinated to a fundamental intentionalism.

B. Joseph Story's Plain-Meaning Purposivism

Joseph Story extended Blackstone's and Marshall's rules into what might seem a proto-textualist approach to the Constitution. He opened his discussion by calling for "uniform rules of interpretation."[57] As with Blackstone, he began with the plain meaning rule: "Where the words are plain and clear, and the sense distinct and perfect arising on them, there is generally no necessity to have recourse to other means of interpretation."[58] And in no circumstances could a court defy the "letter of the law."[59] In Story's view, "liberality of exposition is clearly inadmissible, if it extends beyond the just and ordinary sense of the terms."[60]

Story offered a critique of extratextual sources that would ring true to any textualist. "[C]ontemporary interpretation," he wrote, "must be resorted to with much qualification and reserve."[61] Relying on state ratification debates presented numerous risks: that some state conventions subscribed to peculiar interpretations in order "to remove local objections, or to win local favour"; that some "[inferred] limitations and objects, which others would have rejected"; that some were "governed by a temporary interest or excitement," while others favored more deeply held values; and that some interpreted its language "strictly and closely," while others gave it a "large and liberal meaning."[62] He harshly criticized the idea, championed by Jefferson, that an interpreter should "recollect the spirit" of the constitutional debates:[63]

---

[56] Sturges v. Crowninshield, 17 U.S. 122, 202 (1819). *See also* William S. Blatt, *The History of Statutory Interpretation: A Study in Form and Substance*, 6 CARDOZO L. REV. 799, 812 n.73 (1984) (collecting cases).

[57] STORY, *supra* note 39, at 383. *See also id.* at 145 ("The constitution is not to be subject to such fluctuations. It is to have a fixed, uniform, permanent construction. It should be . . . not dependent upon the passions or parties of particular times, but the same yesterday, to-day, and for ever.").

[58] *Id.* at 384.

[59] *Id.* at 386.

[60] *Id.* at 412.

[61] *Id.* at 388.

> Now, who does not see the utter looseness, and incoherence of this canon. How are we to know, what was thought of particular clauses of the constitution at the time of its adoption? . . . Is the constitution of the United States to be the only instrument, which is not to be interpreted by what is written, but by probable guesses, aside from the text? . . . It is obvious, that there can be no security to the people in any constitution of government, if they are not to judge of it by the fair meaning of the words of the text; but the words are to be bent and broken by the "probable meaning" of persons, whom they never knew, and whose opinions, and means of information, may be no better than their own? [sic] The people adopted the constitution according to the words of the text in their reasonable interpretation, and not according to the private interpretation of any particular men.[64]

All of this would seem to make Story a textualist.[65] But Story's seeming textualism is undercut by three key features that mark him as a purposivist. First is a reliance on intention. Like Blackstone, he prefaced his catalog of interpretive guidelines—including the plain meaning rule—with a statement about the preeminence of intention: "The first and fundamental rule in the interpretation of all instruments is, to construe them according to the sense of the terms, and the intention of the parties."[66]

---

[62] *Id.* at 388–89.

[63] *Id.* at 390 n.1.

[64] *Id. See also id.* at 390 ("Contemporary construction is properly resorted to, to illustrate, and confirm the text, to explain a doubtful phrase, or to expound an obscure clause . . . . It can never abrogate the text; it can never fritter away its obvious sense; it can never narrow down its true limitations; it can never enlarge its natural boundaries.").

[65] Robert Spoo, *"No Word Is an Island": Textualism and Aesthetics in Akhil Reed Amar's* The Bill of Rights, 33 U. RICH. L. REV. 537, 552 (1999) (stating that Story "packed much of the textualist position" into the passage accompanying note 64, *supra*).

And even as Story belittled the use of contemporary interpretation and records from the Constitutional Convention and the ratification debates, he welcomed the use of extratextual indicia of meaning generally: "Mr. Justice Blackstone has remarked, that the intention of a law is to be gathered from the words, the context, the subject matter, the effects and consequence, or the reason and spirit of the law."[67] Nor did he provide a hard plain meaning rule: "Where the words [of the Constitution] are plain and clear," he wrote, "and the sense distinct and perfect arising on them, there is *generally* no necessity to have recourse to other means of interpretation."[68] The sort of ambiguity that would warrant consultation of other sources need not arise from the text itself but could stem from a disagreement between the text and extratextual sources:

> It is only, when there is some ambiguity or doubt arising from other sources, that interpretation has its proper office. There may be obscurity . . . from the doubtful character of the words used, from other clauses in the same instrument, or from an incongruity or repugnancy between the words, and the apparent intention derived from the whole structure of the instrument, or its avowed object.[69]

Put this way, the plain meaning rule merely prevents a court from looking beyond the text when it does not clearly say what history or good sense says that it should. This rule would seem to make Story a weak purposivist, not a plain meaning literalist.

Story's willingness to find ambiguity from extratextual sources is linked to a second trademark of his interpretive approach: a purposivist presumption of the constitutionality of federal exercises of power. Story wrote that an interpreter of the Constitution was, "in the first instance, to consider, what are its nature and objects, its scope and design, as apparent from the structure of the instrument, viewed as a whole, and also viewed in its component parts."[70] Story found the Constitution's nature and objects in its Preamble:

---

[66] STORY, *supra* note 39, at 383.

[67] *Id.*

[68] *Id.* at 384 (emphasis added).

[69] *Id.*

> But a constitution of government, founded by the people for themselves and their posterity, and for objects of the most momentous nature, for perpetual union, for the establishment of justice, for the general welfare, and for a perpetuation of the blessings of liberty, necessarily requires, that every interpretation of its powers should have a constant reference to these objects.[71]

Story read these statements of principle from a nationalist perspective.[72] He was a Republican,[73] but of a Jacksonian sort, with an expansive view of federal power.[74] In his conception of government, he closely emulated the Federalist Marshall, his ally and mentor.[75] He rejected the Jeffersonian view, espoused by Tucker, that federal power should be interpreted strictly, with a presumption of unconstitutionality.[76] Instead, he asserted that grants of power to the federal government should be interpreted broadly:

---

[70] *Id.* at 387. *See also id.* at 404 ("The constitution of the United States is to receive a reasonable interpretation of its language, and its powers, keeping in view the objects and purposes, for which those powers were conferred.").

[71] *Id.* at 406.

[72] *See* Powell, *supra* note 40, at 1301 (describing Story's work as a "massive attempt to prove that the doctrines—nationalism, expansive construction of federal power, and judicial supremacy—for which Story stood and which Jefferson had opposed were in fact the logical conclusions of a truly republican faith").

[73] *Id.* at 1290.

[74] *Id.* at 1293.

[75] Christopher L. Eisgruber, *John Marshall's Judicial Rhetoric*, 1996 SUP. CT. REV. 439, 461 (1996). *See also* PAUL W. KAHN, LEGITIMACY AND HISTORY: SELF-GOVERNMENT IN AMERICAN CONSTITUTIONAL THEORY 37 (1992) (placing Marshall and Story on the nationalist side of the states' rights debate and stating that they saw the Constitution not as an interstate contract but as "the work of the people of the entire nation acting in their unified, sovereign capacity").

[76] Caleb Nelson, *Originalism and Interpretive Conventions*, 70 U. CHI. L. REV. 519, 576 & n.229 (2003) (quoting TUCKER, *supra* note 50, at 154) ("the powers delegated to the federal government, are, in all cases, to receive the most strict construction that the instrument will bear, where the rights of a state or of the people, either collectively, or individually, may be drawn in question"). *See also* Powell, *supra* note 40, at 1300 (stating that Jefferson criticized Story as a "pseudo-Republican").

> By a reasonable interpretation, we mean, that in case the words are susceptible of two different senses, the one strict, the other more enlarged, that should be adopted, which is most consonant with the apparent objects and intent of the constitution; that, which will give it efficacy and force, as a *government*, rather than that, which will impair its operations, and reduce it to a state of imbecility.[77]

Story continued:

> [T]he public functionaries must be left at liberty to exercise the powers, with which the people . . . have entrusted them. They must have a wide discretion, as to the choice of means; and the only limitation upon that discretion would seem to be, that the means are appropriate to the end.[78]

Story offered no textual support for this presumption, instead relying on a lengthy—and similarly non-textualist—quote from *Gibbons v. Ogden*[79] and a deduction on the differences between a constitutional democracy and a constitutional monarchy. It made sense, Story reasoned, to construe strictly a constitutional grant of power to a monarch, since such a grant of power was inherently "odious."[80] Democratic delegations of power to a representative government, however, bore no taint of tyranny and warranted no such restriction.[81]

It is not Story's nationalism that rendered him a purposivist; all interpreters subscribe to ideologies of one sort or another. Nor are substantive presumptions necessarily purposivist;[82]

---

[77] STORY, *supra* note 39, at 404.

[78] *Id.* at 417. *See also id.* at 406 ("No interpretation of the words, in which those powers are granted, can be a sound one, which narrows down their ordinary import, so as to defeat [the Framers'] objects. That would be to destroy the spirit, and to cramp the letter.").

[79] *Id.* at 402–03 (citing 22 U.S. 1, 187-89 (1824)).

[80] STORY, *supra* note 39, at 394.

[81] *Id.* at 396–97.

[82] *See* Scalia, *supra* note 10, at 29 ("The rule of lenity is almost as old as the common law itself, so I suppose that is validated by sheer antiquity.").

even a textualist will embrace a presumption that arises from the text of the Constitution or from the founding generation's understanding of common law.[83] But Story offered no textual basis for his constitutionality presumption, and he acknowledged that it deviated from understandings of governmental power under England's constitutional monarchy. By reaching this conclusion from his own deduction and from historical knowledge, not from the constitutional text or pre-constitutional common law, Story's approach was essentially purposivist.

The primacy of Story's purposivist constitutionality presumption relegates the plain meaning rule to a secondary role. Taken out of context, his instruction to follow the "letter of the law"[84] would suggest a text-centered interpretive approach. But for Story, text was not the start of interpretation but the end—the outer boundary beyond which interpretation could not go. Considering Story's purposivist presumptions and the paucity of constitutional clauses with indisputably clear meanings, however, this outer limit is but a weak restriction on judicial interpretation.

C. Following Story's Lead: Plain-Meaning Purposivism to 1868

A series of treatise writers in the 1850s and 1860s, most notably Thomas Cooley, adopted Story's nationalistic, pseudo-textualist approach to the Constitution. They endorsed his purposivist presumption of the constitutionality of federal exercises of power and, to a greater degree than Story, advised restricting the use of extratextual sources.

---

[83] But in general, textualists shun substantive canons of interpretation. Eskridge, *New Textualism*, *supra* note 15, at 663. *See also* Scalia, *supra* note 10, at 27 (referring to substantive presumptions as rules that "load the dice for or against a particular result"); *id.* at 23 ("A text should not be construed strictly, and it should not be construed leniently; it should be construed reasonably, to contain all that it fairly means.").

[84] STORY, *supra* note 39, at 386. *See also id.* at 412.

1. The Surprisingly Aggressive Purposivism of Theodore Sedgwick

In an 1857 treatise, the New York lawyer[85] Theodore Sedgwick adapted Story's model, but with an outright prohibition on extratextual sources. While Story stated that a court should "generally" not use alternative interpretive techniques when faced with a plain meaning,[86] Sedgwick stated that "[i]t seems to be settled in regard to constitutions as to statutes, that no extrinsic evidence can be received as to their intent or meaning."[87]

Sedgwick acknowledged that close adherence to the text came with certain risks, but he asserted that those of "loose and careless" interpretation were still greater.[88] In his view, a judge should not worry if an unwise constitutional provision compelled an unwise result. He quoted with approval a New York judge whose hard-nosed positivism foreshadowed that of Justice Scalia: "'My rule has been to follow the fundamental law as it [is] written, regardless of consequences. If the law does not work well, the people can amend it . . . .'"[89]

But Sedgwick, an intention-based interpreter,[90] followed Story in adopting a strong presumption of the constitutionality of the federal exercise of power. "[I]n cases of doubt every possible presumption and intendment will be made in favor of the constitutionality of the act in question, and that the courts will only interfere in cases of clear and unquestioned violation of the fundamental law."[91] In empowering judges to determine the meaning of that fundamental law, Sedgwick went beyond Story, outlining a broadly purposivist view that approached dynamic interpretation:

---

[85] Paul D. Carrington, *Legal Education for the People: Populism and Civic Virtue*, 43 U. KAN. L. REV. 1, 15 n.124 (1994).

[86] *See supra* text accompanying note 69.

[87] THEODORE SEDGWICK, A TREATISE ON THE RULES WHICH GOVERN THE INTERPRETATION AND APPLICATION OF STATUTORY AND CONSTITUTIONAL LAW 489 (New York, John S. Voorhies 1857) (citing Eakin v. Raub, 1825 WL 1913, *18 (Pa. 1825) (Gibson, J., dissenting)). *But see infra* text accompanying note 129 (stating that Sedgwick contradicted this extratextual source prohibition in a later edition).

[88] SEDGWICK, *supra* note 87, at 493. *But see infra* text accompanying notes 128–29 (stating that this warning was omitted in a later edition).

[89] SEDGWICK, *supra* note 87 at 486 (quoting with approval Oakley v. Aspinwall, 3 Coms. 547, 568 (N.Y. App. 1850) (Bronson, C.J., dissenting)).

[90] SEDGWICK, *supra* note 87, at 227 ("When we see what is the sense that agrees with the intention of the instrument, it is not allowable to wrest the words to a contrary meaning.") (emphasis omitted from original).

> In regard to a statute, the general duty of the judge is that of a subordinate power, to ascertain and to obey the will of a superior; in regard to a constitution, his functions are those of a co-ordinate authority, to ascertain the spirit of the fundamental law, and so to carry it out as to avoid a sacrifice of those interests which it is designed to protect. No absolute rules of interpretation . . . can be framed.[92]

Granting courts such extensive discretion meant suspending the agency model that Sedgwick applied for statutes. In constitutional matters, he wrote, "the judicial function is really merged in the legislative;" they "vest a sort of legislative power in the judge."[93]

The abolitionist[94] writers Timothy Farrar and Joel Tiffany echoed Sedgwick's strong plain meaning rule—if not his aggressive brand of constitutional purposivism—in a pair of treatises published during Reconstruction. Farrar criticized the loose, extratextual approach of the Southern opposition, who endeavored to not only "tone down or fritter away" the meaning of constitutional clauses, but to "construe them out of the Constitution, and get rid of them entirely."[95] Both writers endorsed a hard plain meaning rule. In Farrar's blunt language, the Framers "meant exactly and only what they said . . . if they meant anything different, it can never be legally proved or known by anybody."[96] But like Story and Sedgwick, both writers endorsed a presumption of the constitutionality of federal action, based on their reading of the goals of the founding generation.[97]

---

[91] *Id.* at 482.

[92] *Id.* at 493.

[93] *Id.* at 230-31.

[94] Bret Boyce, *Originalism and the Fourteenth Amendment,* 33 WAKE FOREST L. REV. 909, 964, 966 (1998).

[95] TIMOTHY FARRAR, MANUAL OF THE CONSTITUTION OF THE UNITED STATES OF AMERICA 127 (Boston, Little, Brown, & Co. 1867). *See also* Boyce, *supra* note 94, at 965 (stating that Tiffany saw the reliance on historical sources in constitutional interpretation as a threat to liberty).

[96] FARRAR, *supra* note 95, at 46. *See also* JOEL TIFFANY, A TREATISE ON GOVERNMENT, AND CONSTITUTIONAL LAW 66 (Albany, N.Y., W.C. Little 1867) ("[O]ne of

2. Thomas Cooley: The Progressive Version

The most notable follower of Story's nationalist, plain-meaning purposivism was Thomas Cooley. By placing him in this tradition, I am deviating from the views of Professors Paul Kahn and Stephen Siegel, who have argued that Cooley followed the non-originalist, history-focused, evolutionary model of constitutional jurisprudence popularized in the late nineteenth century.[98] While Cooley may have been more sensitive to history and custom than Story or Sedgwick, he still subscribed to Story's nationalist pseudo-textualism, unlike his more radical successors. In my view, this paradigm provides a more valuable—and accurate—context for Cooley's work than his use of history.

Cooley expounded on the plain meaning rule in his 1868 treatise, *Constitutional Limitations*. Like his predecessors, he wrote that the meaning of the Constitution was communicated by its plain language[99] and that a court should avoid extratextual sources.[100] Cooley's rule on extratextual sources followed the harder Sedgwick/Farrar/Tiffany plain meaning rule, not the soft Story version; he stated that a court was justified in

---

the first maxims in respect to interpretation is, that it is not allowable where there is no uncertainty as to the meaning of the language used.").

[97] FARRAR, *supra* note 95, at 76-77 (stating that it was the duty of the government to honor the people's constitutionally expressed desire for a strong national government); TIFFANY, *supra* note 96, at 70-71 (stating that Americans "intended to commit their rights and interests . . . to the absolute protection of the government" and that its powers were to be interpreted liberally).

[98] KAHN, *supra* note 75, at 73; *See also*, Stephen A. Siegel, *Historism in Late Nineteenth-Century Constitutional Thought*, 1990 WIS. L. REV. 1431, 1503–08 (1990) [hereinafter Siegel, *Historism*].

[99] THOMAS M. COOLEY, A TREATISE ON THE CONSTITUTIONAL LIMITATIONS WHICH REST UPON THE LEGISLATIVE POWER OF THE STATES OF THE AMERICAN UNION 58 (Boston, Little, Brown, & Co. 1868) [hereinafter COOLEY, CONSTITUTIONAL LIMITATIONS]. *See also id.* at 57 (quoting with approval Newell v. People, 7 N.Y. 9, 97 (1852) (stating that courts could not contradict the clear meaning of the New York Constitution)).

[100] COOLEY, CONSTITUTIONAL LIMITATIONS, *supra* note 99, at 65.

consulting external sources "only" in case of ambiguity.[101] "We are not to import difficulties into a constitution," Cooley wrote, "by a consideration of extrinsic facts, when none appear upon its face.[102]

Also, like his predecessors, Cooley held an intention-based view of the Constitution's meaning: "The object of construction, as applied to a written constitution, is *to give effect to the intent of the people in adopting it*. . . . [I]t is the intent of the lawgiver that is to be enforced. But this intent is to be found in the instrument itself."[103] In embracing an intention-based approach, Cooley flatly rejected the evolutionary model that would develop later in the century. He stated that constitutions were "fixed"[104]—again contradicting the arguments of Professors Kahn and Siegel:

> A cardinal rule in dealing with written instruments is that they shall receive an unvarying interpretation, and that their practical construction is to be uniform. A constitution is not to be made to mean one thing at one time, and another at some subsequent time when the circumstances may have so changed as perhaps to make a different rule in the case seem desirable.[105]

It was not until the emergence of evolutionary constitutionalism that treatise writers would reject original intent, as I discuss in Part IV.

Like Story, Cooley held a strong faith in government power. This faith served his progressive politics—despite the common perception today that he was a conservative.[106] Cooley was troubled by the Gilded Age centralization of wealth[107] and questioned the ability of the private sector to ensure the common welfare.[108] An early defender of the administrative state,[109] he believed that grants of power had to be flexible enough to enable government to respond to changing circumstances.[110]

---

[101] *Id.*

[102] *Id.*

[103] *Id. at* 55. *See also id.* at 57 (stating that courts should seek to determine legislators' intent); Thomas M. Cooley, *The Federal Supreme Court—Its Place in the American Constitutional System*, *in* CONSTITUTIONAL HISTORY OF THE UNITED STATES AS SEEN IN THE DEVELOPMENT OF AMERICAN LAW 31 (New York, G.P. Putnam's Sons 1889) [hereinafter Cooley, *Supreme Court*] (stating that "national authority" was "conferred and measured exclusively by the written instrument").

[104] COOLEY, CONSTITUTIONAL LIMITATIONS, *supra* note 99, at 55.

[105] *Id.* at 54. *See also* Adam Winkler, *A Revolution Too Soon: Woman Suffragists and the "Living Constitution"*, 76 N.Y.U. L. REV. 1456, 1465 n.39 (2001) (citing COOLEY, CONSTITUTIONAL LIMITATIONS, *supra* note 99, at 54-55) (observing that Cooley rejected the notion of an evolutionary Constitution).

It is fitting, therefore, that Cooley endorsed Story's broad, purposivist presumption of federal power. In his view, government exercises of power could be rendered unconstitutional only by express constitutional language, never by mere implication:

> But it is only in express constitutional provisions, limiting legislative power and controlling the temporary will of a majority, by a permanent and paramount law, settled by the deliberate wisdom of the nation, that I can find a safe and solid ground for the authority of courts of justice to declare void any legislative enactment.[111]

In his view, government exercise of police power was rarely unconstitutional, provided it was not "pushed to an extreme that [would] deny just liberty."[112] He saw the role of judge as less that of a critic or a check on power than that of a broad-minded statesman.[113] Tellingly, in Cooley's twenty years as a justice on the Michigan Supreme Court, only once did his court provoke genuine controversy by overturning legislation.[114]

---

[106] *See* FRIEDMAN, *supra* note 3, at 480 ("Cooley himself was much less a cardboard reactionary than has been pictured.").

[107] David M. Rabban, *The Historiography of Late Nineteenth-Century American Legal History*, 4 THEORETICAL INQ. L. 541, 569 (2003).

[108] Paul D. Carrington, *The Constitutional Law Scholarship of Thomas McIntyre Cooley*, 41 AM. J. LEGAL HIST. 368, 370 (1997). *See also id.* at 395 (stating that Cooley's progressivism would have put him on the dissenting side of the *Lochner* Court);

[109] KAHN, *supra* note 75, at 75.

[110] Siegel, *Historism*, *supra* note 98, at 1494.

[111] COOLEY, CONSTITUTIONAL LIMITATIONS, *supra* note 99, at 172. Cooley primarily addressed state constitutions, but the categorical nature of the rules suggests that they apply to the federal constitution as well. *See generally id.* at 38–84. *See also id.* at 171 (stating that a court could not limit federal exercises of power based on a judge's assessment of the "spirit" of the federal Constitution).

[112] THOMAS M. COOLEY, THE GENERAL PRINCIPLES OF CONSTITUTIONAL LAW IN THE UNITED STATES OF AMERICA 227 (Boston, Little, Brown, & Co. 1880).

D. The Last of the Story Disciples?

With few exceptions, treatise writers after 1868 abandoned the plain meaning rule.[115] The Boston jurist[116] Joel Prentiss Bishop gave a half-hearted, contradictory rendering of Story's rules in an 1886 treatise, continuing the purposivist focus[117] but with a weak constitutionality presumption.[118] He repeated the plain meaning rule,[119] only to undercut it by stating that words generally have multiple meanings[120] and that "a court . . . is not required to stultify itself; but it may take into the account any pertinent matter whereof it has judicial cognizance."[121]

As Professor Siegel has written, Bishop the judge subscribed to a principles-based jurisprudence,[122] often failing to articulate or credit the principles on which he was relying.[123] Bishop's contradictory rules on interpretation suggest that he was reluctant to put aside the teachings of Joseph Story but had already been persuaded by the conventional purposivism that would emerge in the late nineteenth century. I discuss this school of thought below in Part III.

---

[113] Cooley, *Supreme Court*, *supra* note 103, at 44 ("A mere lawyer might see in the Constitution nothing but an agreement of parties, to be construed by technical rules; it required a statesman to understand its full significance, as an instrument of government instinct with life and with authority.").

[114] Paul D. Carrington, *Law as the Common Thoughts of Men": The Law-Teaching and Judging of Thomas McIntyre Cooley*, 46 STAN. L. REV. 497, 540 (1997).

[115] *But see* Platt Potter, *Of Constitutional Interpretation*, *in* FORTUNATUS DWARRIS, A GENERAL TREATISE ON STATUTES: THEIR RULES OF CONSTRUCTION, AND THE PROPER BOUNDARIES OF LEGISLATIVE AND JUDICIAL INTERPRETATION 655-56 (Platt Potter ed., Albany, N.Y., William Gould & Sons 1871) (quoting with approval Story's articulation of the plain meaning rule).

[116] Stephen A. Siegel, *Joel Bishop's Orthodoxy*, 13 LAW & HIST. REV. 215, 218 (1995) [hereinafter Siegel, *Bishop's Orthodoxy*].

[117] JOEL PRENTISS BISHOP, COMMENTARIES ON THE LAW OF STATUTORY CRIMES 81–82 (2d ed., Boston, Little, Brown, & Co.1883) ("The rule that the intent of the makers . . . shall prevail over the latent meanings of words and phrases, but not to the disregard of the true import of what is plain, is applied to constitutions the same as to statutes.") (footnote omitted).

[118] *Id.* at 80 ("No mere general considerations will authorize a court to nullify a statute as unconstitutional.").

[119] *Id.* at 58 ("Where the legislative meaning is plain, there is, not only no occasion for rules to aid the interpretation, but it is contrary to the rules to employ them.").

[120] *Id.* at 83.

[121] *Id.* at 59 (discussing statutory interpretation). *See also id.* ("Constitutions are generally to be interpreted by the same rules as statutes").

[122] *See* Siegel, *Bishop's Orthodoxy*, *supra* note 116, at 230.

E. A Not-Quite-So-Plain Meaning School

The plain meaning school has taken a beating in legal commentary. They are apt to be dismissed as natural-law devotees or thick-skulled formalists, unable or unwilling to consider the policy context of the laws they interpret.[124] While this view is no doubt true of many court decisions (but not, I would suggest, *Dred Scott*, which was an example less of mechanistic formalism than of aggressive pro-slavery activism), it hardly does justice to the Story school's intellectual inclinations. They interpreted the Constitution based on their understanding of its principles and history, not on natural law assumptions or sterile, formalist paradigms. They were not textualists or originalists of the Scalia variety. Story and his disciples cleaved to the plain meaning rule, but only as a tactic to be used after arriving at a proper—and broadly purposivist—understanding of the Constitution's principles.

But if Story and his followers are innocent of the charges of wooden formalism or natural-law naïveté, they still deserve criticism for crafting an interpretive approach founded on a double standard. They belittled the value of extratextual sources and discouraged their readers from consulting them, but they failed to acknowledge that their own understandings were based on an inquiry into sources beyond the Constitution's text. They passed the conclusions of this inquiry along to their readers in the form of substantive rules presented as uncontroversial maxims of interpretation. In doing so, they invited their readers to play a pseudo-textualist game, fixed to ensure that their nationalist team would always win.

---

[123] *Id.* at 231. In this way, Bishop perhaps showed a true fidelity to Story. *See infra* Part II.E.

[124] *E.g.*, Molot, *supra* note 35, at 47.

III. A MORE DIRECT PURPOSIVISM: THE RETREAT OF THE PLAIN MEANING RULE

> Perhaps the safest rule of interpretation is to look to the nature and objects of the particular powers, duties, and rights, with all the lights and aids of contemporary history, and to give to the words of each clause such operation and force as is consistent with its legitimate meaning, and not nullify or evade them by astute verbal criticism without regard to the aim and objects of the instrument, and the principles upon which it is based.
>
> —*Robert Desty*[125]

The latter third of the nineteenth century saw a series of treatise writers whose approach to the Constitution resembled modern-day purposivism. Like Story, these theorists based their interpretations not primarily on the Constitution's text but on their understanding of its principles and purposes. But they removed Story's disingenuous plain meaning rule from their interpretive toolboxes and openly embraced a purposivist approach. I call these theorists conventional constitutional purposivists.

Theodore Sedgwick's changing viewpoint provides the most vivid illustration of the disappearance of the plain meaning rule. Sedgwick, an early adopter of the hard plain meaning rule, stated in his 1857 treatise that it was "settled" that a court could not consult "extrinsic evidence" in constitutional interpretation.[126] By the publication of his 1874 edition, he had reversed course, stating that "[i]t is well settled that aid, in regard to the construction of the Constitution of the United States, *may* be derived from contemporaneous exposition and legislative exposition."[127] Similarly, his 1857 edition had included a suggestion that the dangers of strict textualism were less severe than those posed by "loose and careless" readings.[128] By 1874, this warning was stricken from Sedgwick's volume entirely.[129]

---

[125] THE CONSTITUTION OF THE UNITED STATES 39 (Robert Desty ed., 2d ed., San Francisco, Sumner Whitney & Co. 1884).

[126] *See supra* text accompanying note 87.

A. *Early Conventional Constitutional Purposivism*

A few treatise writers in the early 1800s deviated from the plain-meaning approach to constitutional interpretation espoused by Story. William Rawle offered no restriction on extratextual sources in his 1825 treatise, stating only that a court should "deduce the meaning from [the Constitution's] known intention and its entire text, and to give effect, if possible, to every part of it, consistently with the unity, and the harmony of the whole."[130] Story's colleague on the Court, Henry Baldwin, eschewed the plain meaning rule and reflected a weak purposivism: "It is proper to take a view of the literal meaning of the words to be expounded, of their connection with other words, and of the general objects to be accomplished by the prohibitory clause, or by the grant of power."[131] But neither of these treatises appears to have been greeted as a substantial contribution to the interpretive discourse.[132]

---

[127] THEODORE SEDGWICK, A TREATISE ON THE RULES WHICH GOVERN THE INTERPRETATION AND CONSTRUCTION OF STATUTORY AND CONSTITUTIONAL LAW 552 (2d ed., New York, Baker, Voorhis & Co. 1874) (emphasis added).

[128] *See supra* text accompanying note 88.

[129] *See* SEDGWICK, *supra* note 127, at 555-56.

[130] WILLIAM RAWLE, A VIEW OF THE CONSTITUTION OF THE UNITED STATES OF AMERICA 28 (Philadelphia, H.C. Carey & I. Lea 1825).

[131] HENRY BALDWIN, A GENERAL VIEW OF THE ORIGIN AND NATURE OF THE CONSTITUTION AND GOVERNMENT OF THE UNITED STATES 10 (Philadelphia, John C. Clarke 1837).

[132] Few subsequent writers referred to Rawle's treatise. An exception was Joel Parker, who referred to it as a work "of a more general and speculative character." JOEL PARKER, CONSTITUTIONAL LAW: WITH REFERENCE TO THE PRESENT CONDITION OF THE UNITED STATES 8 (Cambridge, Welch, Bigelow, & Co. 1862). Baldwin's appears to have been hampered by its scattershot quality, which perhaps resulted from his deteriorating mental condition. Earl M. Maltz, *Majority, Concurrence, and Dissent:* Prigg v. Pennsylvania *and the Structure of Supreme Court Decisionmaking*, 31 RUTGERS L.J. 345, 374 (2000).

Francis Lieber provided a more intellectually robust post-plain-meaning approach in his treatise, *Legal and Political Hermeneutics*. Lieber, a German émigré to the United States,[133] offered a Burkean model of jurisprudence,[134] where a judge would be guided not only by abstract philosophical principles but by culture, history, and the current needs of the country.[135] "[H]e must fairly interpret according to the spirit of the law and constitution, of constitutional and representative liberty in general, and the genius and history of his people in particular."[136] A judge must interpret the Constitution according to "the true spirit, which pervades the whole constitution . . . provided this spirit is in favor of public welfare . . . or the instrument be not irreconcilable with the present time."[137]

Unlike Story, whom he counted as a friend,[138] Lieber saw little in the constitutional text that was plain, asserting instead that it always required judicial interpretation.[139] And when the text conflicted with the spirit of the law, it was the spirit that must prevail, in strongly purposivist fashion.[140] In rebutting literalist views of the Constitution, Lieber warned his readers "not to hang burdens of great weight upon slight pegs; for instance, an argument of the highest national importance upon the casual position of a word."[141] Lieber named this broad, boldly purposivist brand of interpretation "[t]ranscendent construction."[142] He acknowledged that this approach to the Constitution carried the potential for abuse[143] but contended that it was necessary in order to protect the values and welfare of the people being governed.[144]

---

[133] Kunal M. Parker, *Context in History and Law: A Study of the Late Nineteenth-Century American Jurisprudence of Custom*, 24 LAW & HIST. REV. 473, 491 (2006).

[134] *Id.* at 491.

[135] Paul D. Carrington, *William Gardiner Hammond and the Lieber Revival*, 16 CARDOZO L. REV. 2135, 2138 (1995) [hereinafter Carrington, *William Gardiner Hammond*]. *See also* 2 FRANCIS LIEBER, MANUAL OF POLITICAL ETHICS: DESIGNED CHIEFLY FOR THE USE OF COLLEGES AND STUDENTS AT LAW 582 (Boston, Charles C. Little & James Brown 1839) [hereinafter LIEBER, POLITICAL ETHICS] (stating that the two fundamental assets for an interpreter of the Constitution were good faith and common sense).

[136] LIEBER, POLITICAL ETHICS, *supra* note 135, at 582.

[137] FRANCIS LIEBER, LEGAL AND POLITICAL HERMENEUTICS, OR PRINCIPLES OF INTERPRETATION AND CONSTRUCTION IN LAW AND POLITICS 186 (Boston, Charles C. Little & James Brown 1839) [hereinafter LIEBER, POLITICAL HERMENEUTICS].

[138] Carrington, *William Gardiner Hammond*, *supra* note 135, at 2137-38.

[139] LIEBER, POLITICAL HERMENEUTICS, *supra* note 137, at vii ("For it seems evident that mathematics alone can wholly dispense with interpretation and construction of some sort . . . .").

[140] LIEBER, POLITICAL ETHICS, *supra* note 135, at 583.

[141] LIEBER, POLITICAL HERMENEUTICS, *supra* note 137, at 178.

Lieber also differed from Story by endorsing a strict construction on government power,[145] which he explained in typically colorful language:

> [L]et, with a manly nation, every thing that is in favor of power, be closely construed; every thing in favor of the security of the citizen and the protection of the individual, comprehensively, for the simple reason, that power is power, and, therefore, able to take care of itself as well, as tending, by its nature, to increase, while the citizen wants protection. For the same reason ought we always to be ready to construe comprehensively in favor of the independence of the judiciary . . . .[146]

Professor Lawrence Lessig has concluded from the interest shown by other scholars in Lieber's work that his agenda was not quite as radical at the time as it may appear now.[147] But this reading overlooks the selective use that mid-century theorists made of Lieber's work. They amply cited *Hermeneutics* for its less controversial material, such as a distinction between the definitions of interpretation and construction.[148] But they generally ignored or dismissed his more innovative propositions. For example, Sedgwick, despite ascribing to an aggressive purposivism that vested judges with quasi-legislative powers,[149] described Lieber's notion of transcendent construction as an idea that "amused the fancy and exhausted the arguments of text writers" but that had little "value for the student of jurisprudence."[150] Lieber's vision would eventually find a receptive audience in the last decades of the nineteenth century,[151] but it failed to stimulate any outpouring of conventional purposivism among the Story-dominated treatise writers of the mid-1800s.

---

[142] *Id.* at 185.

[143] *Id.*

[144] *Id.* at 186.

[145] *Id.* at 183–84.

[146] *Id.* at 188.

[147] Lawrence Lessig, *The Limits of Lieber*, 16 CARDOZO L. REV. 2249, 2259 (1995) (stating that "[t]he dominance which Lieber's work quickly gained in interpretive accounts of the time suggests it was not so radical" and lamenting the supposed abandonment of Lieber's views by subsequent courts).

[148] SEDGWICK, *supra* note 87, at 288. *See also* COOLEY, CONSTITUTIONAL LIMITATIONS, *supra* note 99, at 44 (citing a rather conventional explanation by Lieber of judicial review).

B. The Next Wave: The Plain Meaning Rule Rescinded

The late-nineteenth-century wave of conventional purposivism began in 1868, the same year that the Fourteenth Amendment was passed. That year, the California[152] Republican[153] John Norton Pomeroy, hailed as one of the leading writers of legal textbooks in the late 1800s,[154] pointed the way toward a conventionally purposivist approach in his *Introduction to the Constitutional Law of the United States.*

Much of what Pomeroy wrote agreed with the Story model. He stated that the Constitution had a fixed meaning[155] and that its text provided an outer limit on interpretation.[156] He hailed the wisdom of the Framers,[157] and he stated that constitutional clauses were to be read broadly[158] and that exercises of federal power were presumably constitutional.[159]

---

[149] *See supra* text accompanying notes 92–93.

[150] SEDGWICK, *supra* note 87, at 227.

[151] *See infra* text accompanying notes 210-15.

[152] AUMANN, *supra* note 3, at 206 n.29.

[153] Norman W. Spaulding, *Constitution as Countermonument: Federalism, Reconstruction, and the Problem of Collective Memory*, 103 COLUM. L. REV. 1992, 2018 n.137 (2003).

[154] AUMANN, *supra* note 3, at 206 n.29.

[155] *See* JOHN NORTON POMEROY, AN INTRODUCTION TO THE CONSTITUTIONAL LAW OF THE UNITED STATES 11 (New York, Hurd & Houghton 1868) ("The Constitution of the United States is peculiar in that it is all written; that it has nothing of tradition. . . . [T]he past is resorted to only for explanation and interpretation of the written word."). *But see* Siegel, *Historism*, *supra* note 98, at 1462 (stating that Pomeroy held an evolutionary view of the Constitution).

[156] POMEROY, *supra* note 155, at 12 (stating that interpretive tools "must be constantly restrained and limited by the letter itself of the written instrument").

[157] *Id.* at 102 (praising the "almost divine prescience" of the Founders).

But unlike Story, Pomeroy offered no restriction on the use of extratextual indicia of meaning, instead reasoning that their use was necessitated by the general nature of the Constitution's language.[160] Pomeroy was fond of historical and other non-legal materials;[161] he advised free use of "[a]ll the aids which the canons of verbal interpretation, or history, or analogies with other forms, or ethics, can contribute."[162] By welcoming the use not only of historical and textual material but contemporary ethics, Pomeroy marked himself as originalist in only the most abstract.

Three years after Pomeroy's book, John Brown Dillon published a set of guidelines indicating an even more aggressive purposivism, where intent could override even the plain meaning of the text: "Words ought to be construed according to the intent of those who use them, and not otherwise."[163] Robert Desty more clearly articulated the strong purposivist position, stating that "[a]dherence to the letter must not be had in opposition to the reason and spirit of the enactment, and to effectuate the object intended, it may be proper to deviate from the usual sense of the words."[164] Similarly, John Randolph Tucker stated that a law could be constitutional if it followed *either* the text or the original intent: "No legislation can be proper which is inconsistent with the letter and spirit of the Constitution . . . ."[165]

---

[158] *Id.* at 13 ("[A]s greater interests are involved which affect the state rather than the individual, all narrow and technical construction should, as far as possible, be avoided; the nature of the writing as an organic law should be allowed its full effect.").

[159] *Id.* at 166. *See also* David J. Barron & Martin S. Lederman, *The Commander in Chief at the Lowest Ebb—A Constitutional History*, 121 HARV. L. REV. 941, 994, 1019-21 (2008) (discussing Pomeroy's expansive view of executive power). *But see* Spaulding, *supra* note 153, at 2019 (describing Pomeroy as a "state-rights nationalist[]").

[160] POMEROY, *supra* note 155, at 13. Pomeroy's interest in distancing himself from Story is suggested by his citation of Story's landmark work but three times in his 1868 textbook. *Id.* at 35 n.1, 36 n.1 , 38 n.2.

[161] Siegel, *Historism*, *supra* note 98, at 1475.

[162] POMEROY, *supra* note 155, at 10.

[163] JOHN B. DILLON, NOTES ON HISTORICAL EVIDENCE IN REFERENCE TO ADVERSE THEORIES OF THE ORIGIN AND NATURE OF THE GOVERNMENT OF THE UNITED STATES OF AMERICA 31 (New York, S.W. Green 1871).

[164] THE CONSTITUTION OF THE UNITED STATES, *supra* note 125, at 40. *See also id.* at 39 (stating that one should interpret the Constitution to "subserve the great objects for which it was made").

Other writers, while not clearly identifiable as strongly or weakly purposivist, reflected a conventionally purposivist orientation by advising interpreters to reason based on broad principle and by not including a restriction on extratextual sources. Theophilus Parsons offered two principles upon which interpretation should be based. First, he stated that the Constitution was founded on the principle that "the utmost liberty [must be] given to the individual."[166] But he then asserted a fundamentally majoritarian view: "The very first principle . . . is, that when men come together to accomplish any great purpose, the will of the majority must govern."[167] Thomas Powell[168] and Albert Putney[169] took a similar principles-based approach.

Like contemporary purposivism, this was not a consistent school of thought but a collection of ideologically diverse thinkers, united by little more than their endorsement of a common interpretive method. Dillon, reflecting a typical late-nineteenth-century progressivism, looked to the "active moral forces which flow from public virtue, popular education, and

---

[165] JOHN RANDOLPH TUCKER, 1 THE CONSTITUTION OF THE UNITED STATES: A CRITICAL DISCUSSION OF ITS GENESIS, DEVELOPMENT, AND INTERPRETATION 372 (Henry St. George Tucker ed., Chicago, Callaghan & Co. 1899). *See also id.* at 365 ("The interpretation must neither strain to . . . diminish the powers of Congress, but by fair reason reach the mind of the Constitution of the United States.").

[166] THEOPHILUS PARSONS, THE POLITICAL, PERSONAL, AND PROPERTY RIGHTS OF A CITIZEN OF THE UNITED STATES: HOW TO EXERCISE AND HOW TO PRESERVE THEM 40 (Cincinnati, National Publishing Co. 1876).

[167] *Id.* at 39.

[168] THOMAS W. POWELL, ANALYSIS OF AMERICAN LAW 64 (2d. ed., Philadelphia, J.B. Lippincott & Co. 1878) (stating that an interpreter's first two duties were to consider the original mischief that prompted a statute's enactment and to deduce the intention of the lawmaker). *See also id.* at 98 (stating that the same rules apply equally to statutes and constitutions).

[169] ALBERT H. PUTNEY, 2 UNITED STATES CONSTITUTIONAL LAW: STATE CONSTITUTIONS, STATUTORY CONSTRUCTION 279 (1908) (stating that the first principle of interpretation was to follow the intention of lawmakers); *id.* at 302 (advising the use of legislative history without including a plain-meaning restriction). Putney's rules refer to statutes, but the context and the chapter heading indicate that they apply to constitutions as well. *Id.* at 275.

Christian principles."[170] The Virginian Tucker, writing at the end of the century, offered a purposivist defense of the constitutionality presumption[171]—a position at odds with the fervent states' rights advocacy he had practiced throughout his career.[172] Putney aimed his expansive purposivism at the protections guaranteed by the Bill of Rights.[173]

This diverse flowering of purposivist thought challenges the view, put forth by Professor Thomas Peebles, that the *Dred Scott* decision prompted a wave of historist, anti-originalist constitutional thought.[174] Conventional purposivism was no doubt a departure from the plain meaning approach, but not a radical one; it merely suspended the plain meaning rule and created a more forthright purposivism than that of Story and his successors. Nor can *Dred Scott* be safely identified as the primary trigger for interpretive change; that decision was handed down in 1856, but the late-nineteenth century wave of conventional purposivism did not begin until twelve years later. The next wave of interpretation, evolutionary constitutionalism, would not begin in earnest until the 1880s, after the intervening events of the Civil War and Reconstruction.[175] The political backlash that followed *Dred Scott*[176] appears not to have been paralleled by any similar reaction in interpretive theory; if the new purposivism was linked directly to a historical event, it was more likely the war, Reconstruction, or the passage of the Reconstruction amendments.

For our contemporary purposes, the most important conclusion to be drawn from this interpretive tendency is simply that purposivism as we know it is much older than we might think. It first appeared in the 1820s and 1830s, and its post-Cooley revival had flourished in interpretive theory of the Constitution for over two decades before the Supreme Court decided *Holy Trinity.*

---

[170] DILLON, *supra* note 163, at 123.

[171] TUCKER, *supra* note 165, at 377-78.

[172] Paul D. Carrington, *Lawyers Amid the Redemption of the South*, 5 ROGER WILLIAMS U. L. REV. 41, 63 (1999) (describing Tucker as the "premier spokesman for states' rights as a principle of constitutional law").

[173] PUTNEY, *supra* note 169, at 160-61 ("A liberal and elastic interpretation must thus be given the United States Bill of Rights; one which will tend to secure substantial rights of liberty and safety to the individual, rather than one based on an exact adherence to the technical meaning generally accorded to its terms.").

[174] Thomas H. Peebles, *A Call to High Debate: The Organic Constitution in Its Formative Era, 1890-1920*, 52 U. COLO. L. REV. 49, 65 (1980).

[175] *See infra* text accompanying note 216.

[176] *See, e.g.*, *infra* text accompanying note 199.

IV. EVOLUTIONARY CONSTITUTIONALISM

> [I]t is the people of the present day who possess the political power, and whose commands give life to what otherwise is a dead letter. No people are ruled by dead men, or by the utterances of dead men. Those utterances are only law so far as they are voiced by some living power. . . . For the present possessors of political power, and not their predecessors, are the lawgivers for the present generation.
>
> —*Christopher Tiedeman*[177]

For most of the nineteenth century, interpretive theory was not anti-originalist. The two theories I have discussed were non-textualist, but they cleaved to originalism, at least in a broad, conceptual sense. If the story ended there, there could be some truth to Justice Scalia's 1989 statement that "[i]t is only in relatively recent years . . . that non-originalist exegesis has, so to speak, come out of the closet, and put itself forward overtly as an intellectually legitimate device."[178]

But Scalia's assertion ignores the late-nineteenth-century movement that arguably represented the most radical major interpretive theory of our history: evolutionary constitutionalism.[179] By abolishing the search for intent, this school sharply broke with both plain-meaning purposivism and conventional purposivism. Like today's living constitutionalists, the evolutionists viewed constitutional law as dynamic, determined not necessarily by the original intent of the founders but by the changing values and circumstances of the people being governed. But unlike living constitutionalists generally, this school conceived of law as the product of evolutionary progression. Their almost mystical notion of law was founded on the notion that the common law evolved through judicial interpretation to keep pace with the collective will and culture of a living, changing nation.

---

[177] CHRISTOPHER G. TIEDEMAN, THE UNWRITTEN CONSTITUTION OF THE UNITED STATES: A PHILOSOPHICAL INQUIRY INTO THE FUNDAMENTALS OF AMERICAN CONSTITUTIONAL LAW 150 (New York, G.P. Putnam's Sons 1890).

[178] Antonin Scalia, *Originalism: The Lesser Evil*, 57 U. CIN. L. REV. 849, 852 (1989).

[179] Rabban, *supra* note 107, at 556 (stating that the term "evolutionary jurisprudence" is used interchangeable with three different schools of thought: historicism, legal evolution, and the Darwinian theory of law).

Because they distanced themselves from—and even scorned—intention-based interpretation, the evolutionists crafted what may be the first anti-purposivist approach to American constitutional interpretation. But because they rejected also any notion of adhering to the text, they must also be thought of as fundamentally anti-textualist. This anti-purposivist, anti-textualist approach gave the evolutionists a powerful tool to address the late nineteenth century's primary legal challenge: defining substantive property rights and the extent of police power[180] in a climate of large-scale commerce and the regulatory state.[181]

A. Sources

1. The Evolutionary Principle

Commentators on the evolutionary school have provided a rich historical view of the intellectual roots of the evolutionary constitutionalists. Perhaps the first lesson to be learned is that the notion of evolution was not the brainchild of Darwin, whose *Origin of the Species* was not published until 1859, but of a series of thinkers[182] who had proposed evolutionary models for science, society, and government dating back to the eighteenth century.

First was Edmund Burke, who, in the eighteenth century, wrote of the state not as a static structure but a growing organism.[183] German romantics built a similar concept around the concept of the *volk*, which figured a country's population as a transcendent, organic entity.[184] Friederich Karl von Savigny applied this idea to law in the early 1800s.[185] In his view, law was an evolving extension of the *volkgeist*—the dynamic, creative spirit of a people—and not a mere governmental creation.[186] This idea became the basis for Savigny's historical school of jurisprudence, which proposed to interpret law not through abstract analysis but through a sensitive study of a nation's cultural and historical foundations.[187]

---

[180] Stephen A. Siegel, Lochner *Era Jurisprudence and the American Constitutional Tradition*, 70 N.C. L. REV. 1, 8 (1992) [hereinafter Siegel, Lochner *Era Jurisprudence*].

[181] *Id.* at 7-8.

[182] *See* JAMES E. HERGET, AMERICAN JURISPRUDENCE, 1870-1970: A HISTORY 145 (1990) (stating that evolution is not a single theory but a family of theories).

Herbert Spencer and Henry Sumner Maine connected the evolutionary model to economic concepts. In his landmark *Social Statics*, published in 1851, Spencer portrayed evolution as a zero-sum game in which players allocated finite resources according to the laws of natural selection.[188] The process should, in his view, be allowed to run its natural course,[189] with a minimum of state interference.[190] Private property rights were a product of this process, an evolutionary improvement over the backwards communal ownership model of previous epochs.[191] Henry Sumner Maine held a similar view. In his conception, legal systems inexorably evolved from basic societies, where rights were based on family or tribal status, to advanced systems in which they were defined by contract.[192]

2. The Anti-Constitutionalists

Evolutionary constitutionalism developed in a post-Civil War age in which the Constitution itself was regarded with suspicion, if not outright scorn. Perhaps the best example of the latter is the Unionist lawyer Sidney George Fisher,[193] whose 1862 polemic, *The Trial of the Constitution*, rebutted the claim that Lincoln's exercise of executive power violated the Constitution.[194] Fisher responded to such claims by criticizing the Constitution itself. He saw it as an inherently flawed document,[195] all but impossible to amend,[196] whose failings were starkly demonstrated by its failure to protect the country from the disastrous Civil War.[197] He criticized what he saw as the reactionary judiciary[198] whose myopic originalism was responsible for the disastrous *Dred Scott* decision.[199]

---

[183] Peebles, *supra* note 174, at 55, 57–58.

[184] *Id.* at 58.

[185] E. Donald Elliott, *The Evolutionary Tradition in Jurisprudence*, 85 COLUM. L. REV. 38, 41 (1985).

[186] *Id.* at 42; Peebles, *supra* note 174, at 58.

[187] Elliott, *supra* note 185, at 40.

[188] Herbert Hovenkamp, *Evolutionary Models in Jurisprudence*, 64 TEX. L. REV. 645, 666-67 (1985) [hereinafter Hovenkamp, *Evolutionary Models*].

[189] *Id.* at 671.

[190] *Id.* at 667.

[191] *Id.* at 668.

[192] AUMANN, *supra* note 3, at 198.

Fisher took aim at the very notion of intent. To Fisher, law resided not in legal instruments but in custom,[200] which he saw as an expression of rational growth.[201] He challenged the notion of a Constitution with a fixed meaning. In his view, "[a] fixed, unchangeable Government, for a changeable, advancing people, is impossible . . . ."[202] The Constitution would have to change or be discarded:

> But, that it may endure, it must be capable of alteration, when an alteration is needed, otherwise it will be destroyed; for a Constitution that cannot be changed for the people, and by the people, must . . . be swept away by the force of opinion, and the march of events.[203]

---

[193] KAHN, *supra* note 75, at 72.

[194] SIDNEY GEORGE FISHER, THE TRIAL OF THE CONSTITUTION 96 (Philadelphia, J.B. Lippincott & Co. 1862).

[195] *Id.* at 20.

[196] *Id.* at 30.

[197] *Id.* at 55. *See also* KAHN, *supra* note 75, at 70 (stating that Fisher viewed the war "as a symbol of the defective character of the founders' Constitution").

[198] Winkler, *supra* note 105, at 1471.

[199] KAHN, *supra* note 75, at 72.

[200] *Id.*

[201] *Id.*

[202] FISHER, *supra* note 194, at 39. *See also id.* at 2 ("No man or assembly of men is wise as the generations or as time, for time reveals imperfections . . . .").

[203] *Id.* at 96.

Fisher's view was not the fleeting sentiment of a country at war. John Codman Hurd, a former abolitionist,[204] attacked the Constitution and intention-based interpretation in an 1881 monograph, *The Theory of Our National Existence.* Hurd embraced the Burkean idea that constitutional questions had to be settled with regard to the will of the sovereign people, not by reference to abstract ideas in the text.[205] In his view, the Story school's reverence for the Constitution and original intent ignored sovereignty.[206] Hurd excoriated what he saw as "Story's fetish [C]onstitution":[207]

> We can now see that the feeling is akin to that in the Guinea negro's ascription to the toy he has himself made, from sticks, rags, and feathers, of the divinity of uncreated essence, and of power in respect to which he, who patched the pieces together, is a helpless slave. . . . The fact that there must, by the nature of things, by the conditions of all political existence, always be a continuing somebody . . . to make it *law* for anybody, is sedulously veiled from the conscience of each American. He must *believe* that the Constitution came,—from heaven or from hell, it matters not; but came,—and that we all are what we are by it, and without it would be nowhere.
>
> What is this but a fetish?[208]

The work of Fisher and Hurd presaged a revolt against the founding generation. As Paul Kahn has observed, these thinkers saw the Founders "reduced from semidivine fathers to naive children," their Constitution not a work of genius but a fortuitous accident.[209]

---

[204] Robert Meister, *Sojourners and Survivors: Two Logics of Constitutional Protection*, 3 U. CHI. L. SCH. ROUNDTABLE 121, 128–30 (1996).

[205] *See* JOHN CODMAN HURD, THE THEORY OF OUR NATIONAL EXISTENCE: AS SHOWN BY THE ACTION OF THE GOVERNMENT OF THE UNITED STATES SINCE 1861, at xviii (Boston, Little, Brown, & Co. 1881) (stating that law depended on "a human agency, a human intellect interpreting the meaning of the words, and a human will and force compelling obedience to that meaning").

[206] *Id.* at ix (criticizing Story and his followers for "seem[ing] to ignore the fact that, in every country in the world, there must be somebody in existence whose measure of right everybody else must accept as the rule of action before there can be any public law at all"). *See also id.* at 109 n.3 (criticizing Story's interpretation of the phrase, "We the people," and stating, "The question is not of *interpretation* of the word *people*, as a *legal* question. It is to identify those who used the pronoun WE, as a *political* question.").

[207] *Id.* at 403.

[208] *Id.* at 95.

3. The Lieber Revival

The late nineteenth century saw a revival of the work of Francis Lieber and, in 1880, a posthumous reissue of his *Hermeneutics*.[210] Lieber's more radical ideas about interpretation had generated little traction among mid-century treatise writers,[211] but they found a receptive audience among late-century scholars. To these revivalists, Lieber's work was a much-needed antidote to the influence of Christopher Columbus Langdell, whose influence threatened to transform the legal profession into a narrow, technocratic vocation divorced from political, social, or economic concerns.[212]

Lieber's interpretive framework was well-suited for an evolutionary conception of law. His view that a judge must interpret not only according to the spirit of the law but to "the genius and history of his people"[213] went well beyond conventional purposivism. His rule that no law could be constitutional when it contradicted the public welfare[214] suggested a jurisprudence more attuned to public policy than to legal abstractions. While Lieber was no anti-constitutionalist, he had articulated a notion of non-originalist sovereignty decades before Fisher or Hurd:

> [C]onstitutions do not make liberty; liberty is not decreed in so many words on parchment. That parchment, with its ink upon it, may be eaten by the worms . . . but if they are but the pronouncing and solemn expression of that which lives within

---

[209] KAHN, *supra* note 75, at 67. *See also id.* (discussing postwar thinkers' "patricidal attitude").

[210] Carrington, *William Gardiner Hammond*, *supra* note 135, at 2143-44.

[211] *See supra* text accompanying notes 147–51.

[212] Carrington, *William Gardiner Hammond*, *supra* note 135, at 2149–50.

[213] LIEBER, POLITICAL ETHICS, *supra* note 135, at 582.

[214] LIEBER, POLITICAL HERMENEUTICS, *supra* note 137, at 181.

> the nation, the written words of the living essence, it is far otherwise.[215]

B. Social Darwinist Constitutionalism: The Conservatives

1. William Gardiner Hammond

From these philosophical sources—evolution, anti-constitutionalism, and the writings of Francis Lieber—the evolutionary constitutionalists crafted a distinctly conservative constitutional ideology. Among the first was a German-educated Iowa lawyer, William Gardiner Hammond, editor of the 1880 edition of Lieber's *Hermeneutics* and perhaps the leading legal historian of his day.[216] Hammond supplemented Lieber's work with a hundred-page appendix articulating the platform of this new school of American legal thought.[217]

Hammond's criticism of the plain meaning school anticipated the one Hurd would make a year later. He described Story's *Commentaries* and Tucker's edition of Blackstone as typical of American legal treatises that purported to interpret law without inquiring into the rationale behind it.[218] In his view, the work of Savigny had put to rest the notion that a judge could only consult extratextual sources in case of ambiguity.[219]

Following Lieber, Hammond called for judges to have the discretion to apply law flexibly to accord with changing human behavior.[220] This discretion was itself a product of evolution:

---

[215] *Id.* at 179.

[216] Carrington, *William Gardiner Hammond*, *supra* note 135, at 2143-44.

[217] *See generally* William G. Hammond, *Appendix* to FRANCIS LIEBER, LEGAL AND POLITICAL HERMENEUTICS (3d ed., St. Louis, F.H. Thomas & Co. 1880).

[218] *Id.* at 306. In my view, Hammond misread Story, whose theory was indeed based on a careful—if not fully disclosed—consideration of the rationale underlying the Constitution. *See supra* text accompanying note 125.

[219] *Id.* at 243 (quoting an uncited passage by C.G. Bruns) ("It is a gross mistake to suppose that the words alone compose the law, and that the meaning or spirit—the more ideal elements—can only be appealed to when the words themselves are ambiguous. . . . Savigny has successfully refuted this view . . . .").

[220] *Id.* at 284.

> Man begins with narrow, strict, technical rules, formal language, arbitrary definitions; and it is only by centuries of labor . . . that he gradually learns to dispense with them, as clumsy tools, to use abstract ideas with some degree of freedom . . . and to form . . . judgments that will comment themselves at once to the reason and conscience of the community.[221]

To Hammond, a dismissive view of the text was permissible because the real Constitution—the one that mattered—was not even written down:

> No truth can be clearer to the student of history and law than that a written constitution of any value always presupposes the existence of an unwritten one. . . . [T]he constitution as an *objective fact* must exist, before the constitution as an *instrument of evidence* can have any value. The worthlessness of written constitutions that have not unwritten ones beneath and behind them, is one of the most frequently recurring lessons of the nineteenth century.[222]

And what powers did this unwritten constitution have? In the spirit of Spencer and Maine, Hammond asserted that it protected economic rights—even, he suggested, to the extent of contradicting the Fourteenth Amendment. "[A] change of the written constitution," he wrote, "does not abrogate rights of property . . . ."[223]

The conservative school of evolutionary constitutionalism would consistently embrace the views that Hammond articulated: an evolving, unwritten Constitution; disregard for the intent of the Framers; the free use of extratextual sources; broad judicial discretion; and a fervent protection of economic rights.

---

[221] *Id.* at 292.

[222] *Id.* at 308. *See also id.* at 309 ("That constitutional government . . . existed in England and America long before the first reduction of the constitution of any state to writing, is too familiar a matter of history to be proved here."); *id.* at 319 (stating that judicial precedent was evidence of the unwritten law).

[223] *Id.* at 310.

2. Francis Wharton

Francis Wharton reflected these principles in his 1884 *Commentaries on Law.*[224] Wharton took a Christian view of the work of Burke, Savigny, and Spencer.[225] An Episcopal priest,[226] Wharton believed that the evolution of society and law was not merely natural but was part of God's plan.[227] Wharton embraced the view that law was determined not by original intention but by the will of the living sovereign, stating that "no laws that conflict with such conscience, or do not respond to such needs, are operative."[228] But he was no majoritarian positivist. Following the German model, he did not see law so much as a conscious, considered choice but as "the instinctive utterance of the people,"[229] a set of "emanations rather than efforts,"[230] a product of an unconscious "sense of right,"[231] the "silent and spontaneous evolution of the nation . . . adapting itself to the conditions in which in each epoch it is placed."[232]

Wharton did not trust elected representatives to convey the substance of this evolving law. "[T]he unconscious tendencies of a Christian nation," he wrote, "are more wise and healthy than the conscious policy of its legislators or rulers, or even of its own component members meeting for popular deliberation."[233] He praised the Framers not for their wisdom or foresight but for giving "substantial expression" to the needs of the nation.[234] For Wharton, only judges could properly articulate the needs and character of the people.[235] Invoking a geologic simile,[236] he saw common law jurisprudence as:

---

[224] *See generally* Stephen A. Siegel, *Francis Wharton's Orthodoxy: God, Historical Jurisprudence, and Classical Legal Thought*, 46 AM. J. LEGAL HIST. 422 (2004) [hereinafter Siegel, *Wharton's Orthodoxy*].

[225] KAHN, *supra* note 75, at 82.

[226] Rabban, *supra* note 107, at 569.

[227] Siegel, *Wharton's Orthodoxy*, *supra* note 224, at 432.

[228] FRANCIS WHARTON, COMMENTARIES ON LAW 56 (Philadelphia, Kay & Brother 1884). *See also id.* at 85–86 (criticizing Blackstone's static view of the Constitution).

[229] *Id.* at 433 (describing the words of the Preamble to the Constitution).

[230] *Id.* at iii.

[231] *Id.* at 99.

[232] *Id.* at iv.

[233] *Id.* at 61. *See also id.* at 93 ("[T]o Mr. Spencer, the law of a nation is the product of custom flowing from and through prior generations, its quality and its course modified by their distinctive characteristics . . . .").

> [A] glacier which is congealed and yet flows. . . . [I]t forms a solid and consistent structure; it incorporates in its volume whatever rights belong to the community . . . yet still it moves on, and not only moves but expands. . . . [I]t moves in complex sympathy with the conscience and genius of the people from whom it emanates. Spontaneously and instinctively there appears a redress for every wrong, and a retribution for every injury.[237]

This disregard for the mechanics of representative government provided a solid basis for Wharton to provide a capitalist revision of the Fourteenth Amendment. Wharton, a defender of laissez-faire economics,[238] asserted that the amendment was not a menace but rather a bulwark against unwelcome regulation of commerce.[239] This was to be the true, lasting purpose of the amendment, regardless of the racial concerns that underlay its passage:[240]

> [T]he provisions contained in these amendments, bearing distinctively on the negro race, are comparatively ephemeral in their character, while the clause before us is likely to be permanent, and to permeate the whole business system of the Union. Almost all the adjudications in respect to the amend-

---

[234] WHARTON, *supra* note 228, at 410 ("[T]he constitution was evolved in a large measure unconsciously from the conditions in which the community was placed; and it by no means follows, because the formal framers of the constitution succeeded in giving substantial expression to the national conscience and needs, that what they said is equally authoritative with what they did and with what caused their action.").

[235] KAHN, *supra* note 75, at 86. *See also* Siegel, *Wharton's Orthodoxy*, *supra* note 224, at 431 (describing Wharton's vision of law as an evolutionary process guided by judicial precedent).

[236] *See* Siegel, *Wharton's Orthodoxy*, *supra* note 224, at 426 (stating that Wharton was a student of geology).

[237] WHARTON, *supra* note 228, at 99-100.

[238] Siegel, *Wharton's Orthodoxy*, *supra* note 224, at 424.

[239] Herbert Hovenkamp, *The Political Economy of Substantive Due Process*, 40 STAN. L. REV. 379, 396 (1988).

[240] *Id.*

> ment have, heretofore, related to negro rights. But these are now finally settled. . . .[241]

3. Christopher Tiedeman

Christopher Tiedeman, author of the tellingly titled *The Unwritten Constitution*, was the best known and most extensively studied of the American evolutionary constitutionalists. While he was by no means the school's founder,[242] he arguably introduced Social Darwinism to more judges than did Spencer himself.[243]

Like Wharton, Tiedeman perceived the Due Process Clause of the Fifth and Fourteenth Amendments as protecting economic interests from government regulation.[244] But he added yet another layer of laissez-faire protection, a presumption of the *un*constitutionality of state police power legislation:

> Wherever by reasonable construction the constitutional limitation can be made to avoid an unrighteous exercise of police power, that construction will be upheld, notwithstanding the strict letter of the constitution does not prohibit the exercise of such a power. The unwritten law of this country is in the main against the exercise of police power. . . .[245]

Tiedeman also surpassed his predecessors in invoking the nation's supposed racial and cultural heritage: "[T]he English Constitution was not the conscious and voluntary creation of the English people . . . it was an evolution from the simple political principles and formulae of the Teutonic race. . . ."[246]

---

[241] WHARTON, *supra* note 228, at 681.

[242] *But see* KAHN, *supra* note 75, at 77–78 (crediting Tiedeman as the movement's leader).

[243] David E. Bernstein, Lochner *Era Revisionism, Revised:* Lochner *and the Origins of Fundamental Rights Constitutionalism*, 92 GEO. L.J. 1, 8 n.21 (2003).

[244] CHRISTOPHER G. TIEDEMAN, A TREATISE ON THE LIMITATIONS OF POLICE POWER IN THE UNITED STATES 10 (St. Louis, The F.H. Thomas Law Book Co. 1886) (hailing the due process clause as "'[t]he main guaranty of private rights against unjust legislation'") (quoting with approval Bertholf v. O'Reilly, 74 N.Y. 509, 515 (1878) (holding that a state liquor regulation violated the due process clause of the New York constitution))[hereinafter TIEDEMAN, POLICE POWER].

[245] TIEDEMAN, POLICE POWER, *supra* note 244, at 10.

[246] TIEDEMAN, *supra* note 177, at 17.

Those principles and formulae provided the two countries with similar political templates: "The unwritten constitution of the United States, within the broad limitations of the written Constitution, is just as flexible, and yields just as readily to the mutations of public opinion as the unwritten constitution of Great Britain."[247]

But Tiedeman's views on constitutional interpretation were fully continuous with those of Hammond and Wharton. In his view, constitutional authority rested in the conscience of the living, sovereign lawgiver,[248] but it was the job of judges, not legislators, to make such determinations.[249] He warned against revering the founders or treating the Constitution as a "popular idol."[250] He encouraged the free use of extratextual sources[251] and viewed textual adherence as a relic of "earlier stages of development of a system of jurisprudence," when the written word is held in "reverential awe."[252] He drew heavily on Lieber's ideas, particularly the idea that a law was unconstitutional when it conflicted with a people's "prevalent sense of right."[253] And, like Wharton, he viewed the Constitution's primary role as promoting commerce, not preventing oppression. "I do not think," he wrote, "there can be much doubt that the danger of official tyranny has been successfully dissipated in the American constitutional system."[254]

---

[247] *Id.* at 43.

[248] *Id.* at 153 (stating that this notion was consistent with the views of Justice Marshall). *See also id.* at 151 (stating that a judge "need not concern himself so much with the intentions of the framers of the Constitution or statute" but instead should "find out what the possessors of political power now mean by the written word").

[249] Siegel, *Historism*, *supra* note 98, at 1528, 1537.

[250] TIEDEMAN, *supra* note 177, at 21. *See also id.* ("It is the complete harmony of its principles with the political evolution of the nation, which justly challenges our admiration, and not the political acumen of the convention which promulgated it."); *id.* at 153 ("This fallacy in interpretation of laws is the result of holding on to a rule, after a change of circumstances has confused its meaning or made its application misleading; and its retention, after it has ceased to be true, is due to the general acceptance of the groundless doctrine of the social contract. Under this doctrine, as well as under the doctrine of the divine right of kings, the popular conception of law was . . . that it emanated from some power above and beyond us, from God in the one case, and from our ancestors in the other case."); Parker, *supra* note 133, at 512 (writing that Tiedeman displayed a "refreshingly cavalier attitude toward the integrity of textual language").

[251] TIEDEMAN, *supra* note 177, at 147-48 ("[E]very fact or circumstance, surrounding the lawgiver, when the law was promulgated, is required by our rules of interpretation to be inquired into; so that the cardinal rule of interpretation of laws may be said to be, that the intention of the lawgiver, when the law was enacted, must prevail.").

[252] *Id.* at 145.

[253] *Id.* at 149. *See also id.* at 149 n.1 (discussing Hammond's explications of Lieber's interpretive rules).

After Tiedeman, treatise writers continued to express this conservative, social Darwinist view of the Constitution through the end of the century. Francis Newton Thorpe discussed the school of thought in starkly economic terms in his 1898 book *A Constitutional History of the American People.* While Tiedeman's economic conservatism was based on an idealized view of the economic activity of small-scale property owners,[255] Thorpe unabashedly celebrated the Constitution as a tool to promote industrial growth. "The industrial process has been coordinated with the civil," he wrote, "and democracy in America is the result."[256]

But Thorpe was effectively bringing up the rear of his conservative constitutional movement. The weaknesses of Social Darwinism as a legal theory became apparent as soon as it moved from the treatise to the courtroom.[257] Social Darwinist constitutionalism re-legitimated the role of judges,[258] but it gave judges little reason to protect constitutional rights.[259] Its lack of regard for the negative effects of commercial activity revealed the fundamental weakness of the Spencerian idea: the assumption that evolution by natural selection was an inherent good and that government interference was unnatural and dangerous.[260] After 1900, evolutionary constitutional jurisprudence would head in a distinctively progressive direction.

---

[254] *Id.* at 162.

[255] Louise A. Halper, *Christopher G. Tiedeman, 'Laissez-Faire Constitutionalism' and the Dilemmas of Small-Scale Property in the Gilded Age*, 51 OHIO ST. L.J. 1349, 1384 (1990) (stating that Tiedeman's preferred model of property and enterprise was not the large modern corporation but the small-scale freeholder). *See also* James W. Ely, Jr., *The Protection of Contractual Rights: A Tale of Two Constitutional Provisions*, 1 N.Y.U. J. L. & LIBERTY 370, 399 (2005) (stating that Tiedeman's "intense dislike of corporate privilege stood at odds with their dedication to open competition in a free market").

[256] FRANCIS NEWTON THORPE, 1 A CONSTITUTIONAL HISTORY OF THE AMERICAN PEOPLE 1776-1850, at 1 (New York, Harper & Bros. 1898). *See also id.* at 7 (stating that "virtue in a democracy lies close to industry"); *id.* at 26 (praising American "social efficiency" and the increasingly economic orientation of the law).

[257] Siegel, Lochner *Era Jurisprudence*, *supra* note 180, at 100.

[258] *Id.* at 76.

[259] Hovenkamp, *Evolutionary Models*, *supra* note 188, at 669 ("The most fundamental political right for Spencer was the 'right to ignore the State.'") (quoting HERBERT SPENCER, SOCIAL STATICS 208 (London, John Chapman1851)).

C. Reform Darwinist Constitutionalism: The Progressives

While the focus of this article is the nineteenth century, no treatment of evolutionary constitutional thought can end with the social Darwinists. A new school, the reform Darwinists, adapted evolutionary logic to progressive ends in the early twentieth century.[261] These thinkers, notably the administrative law scholar Frank Johnson Goodnow[262] and President-to-be Woodrow Wilson[263]—abandoned the social Darwinist aversion to interfering with the supposed natural course of evolution.[264] Drawing on the pragmatism of John Dewey,[265] the sociological jurisprudence of Roscoe Pound,[266] and the administrative law work of Ernst Freund, [267] this group hailed the human ability to solve social problems as itself a valuable product of evolution.[268] In their view, it should not be suppressed but should be used to structure and even accelerate the evolution of society. [269]

---

[260] Hovenkamp, *Evolutionary Models*, *supra* note 188, at 671.

[261] *Id.* at 671; Peebles, *supra* note 174, at 65.

[262] FRANK J. GOODNOW, SOCIAL REFORM AND THE CONSTITUTION 6 (1911) ("[C]onstitutions . . . which are practically unamendable, should be considered rather as statements of general principles whose detailed application should take account of changing conditions, and should be so interpreted by judicial decisions as to be susceptible of a continuous and uninterrupted development."); Mark Fenster, *The Birth of a "Logical System": Thurman Arnold and the Making of Modern Administrative Law*, 84 OR. L. REV. 69, 77 (2005) (stating that Goodnow was the first professor to teach a law school course in administrative law).

[263] WOODROW WILSON, CONSTITUTIONAL GOVERNMENT IN THE UNITED STATES 192 (1911) ("The Constitution cannot be regarded as a mere legal document, to be read as a will or a contract would be. It must . . . be a vehicle of life. As the life of the nation changes so must the interpretation of the document which contains it change . . . determined, not by the original intention of those who drew the paper, but by the exigencies and the new aspects of life itself."). *See also* Eric R. Claeys, *Progressive Political Theory and Separation of Powers on the Burger and Rehnquist Courts*, 21 CONST. COMMENT. 405, 409 (2004) (stating that Goodnow shared with Wilson an interest in an "independent and centralized national bureaucracy").

[264] Hovenkamp, *Evolutionary Models*, *supra* note 188, at 671.

[265] Peebles, *supra* note 174, at 77.

[266] Hovenkamp, *Evolutionary Models*, *supra* note 188, at 678 (stating that Pound proposed to base constitutional rights not on metaphysical or natural law propositions but on scientifically verifiable public policy claims).

In legal terms, this meant deciding constitutional questions not with regard to the historical values of a race or a nation but according to the principles and needs of the current generation.[270] And it meant a shift from the social Darwinists' expansive view of judicial power to a narrow, restricted one. Reform Darwinists applied the teachings of James Bradley Thayer, whose deferential, majoritarian[271] approach to judicial review provided the basis for what we know as rational basis review.[272] Thayer opposed courts' invocation of property rights to render legislation unconstitutional and consistently defended legislative exercise of the police power. [273] The reform Darwinists proposed using Thayer's model to give legislatures and agencies the freedom they needed to implement reforms—and to give the new discipline of administrative law the room it needed to grow.[274]

---

[267] Fenster, *supra* note 262, at 77-78 (stating that Freund called for increased governmental reliance on disinterested experts and the creation of a new field of administrative law to enable more effective regulation of economic activity).

[268] Hovenkamp, *Evolutionary Models*, *supra* note 188, at 671.

[269] Peebles, *supra* note 174, at 77.

[270] Siegel, Lochner *Era Jurisprudence*, *supra* note 180, at 102.

[271] William Michael Treanor, *Process Theory, Majoritarianism, and the Original Understanding*, 75 FORDHAM L. REV. 2989, 2991 (2007) (describing Thayer's *The Origin and Scope of American Constitutional Law* as the "urtext of the majoritarian movement").

[272] Foreword, *One Hundred Years of Judicial Review: The Thayer Centennial Symposium,* 88 NW. U. L. REV. [i], [i] (1993). *See also* KAHN, *supra* note 75, at 85 (describing Thayer's criticism of judicial review as an accident of history and contrary to the intent of the Framers).

[273] Jay Hook, *A Brief Life of James Bradley Thayer,* 88 NW. U. L. REV. 1, 7 (1993).

[274] *See* ERNST FREUND, THE POLICE POWER: PUBLIC POLICY AND CONSTITUTIONAL RIGHTS v (1904) (stating that the development of administrative law had to be largely "constructive," given the newness of the field). *See generally* W.G. Hastings, *The Development of Law as Illustrated by the Decisions Relating to the Police Power of the State*, 39 PROC. AM. PHIL. SOC'Y 359 (1900) (conceding that the police power was a fiction defined by custom, not the Constitution).

The appearance of reform Darwinism illustrates the often dialectic nature of interpretive theory. One group develops a theory to pursue ideological ends, only to have that theory appropriated by those who would use it for entirely different objectives. The political content changes, but the theoretical form remains the same. In its deference to administrative and legislative decisions, the progressive school was a return to Story and Cooley. But because it maintained the evolutionary logic and the anti-originalism of its conservative predecessor, it remained essentially anti-purposivist—and anti-textualist.

CONCLUSION

This article has provided a history and taxonomy of the major theories of constitutional interpretation in the nineteenth century. By century's end, courts could choose any of three broad approaches to the text. They could apply the ageing Story model, following what they saw as the Constitution's plain meaning when considered from a nationalist, purposivist perspective.[275] They could abandon the plain meaning rule and turn to alternative sources to determine a clause's purpose and meaning, even to the extent of contradicting the text. Or they could dismiss both the text and any evidence of original meaning, considering instead to the character of the country or the current exigencies of government.

Judges of the time probably would not have perceived this orderly menu I have described. In practice, judges are eclectics. They rely on a variety of schools of thought, often discovered indirectly and filtered through their own ideological and philosophical conceptions. But an understanding of the broad trends in constitutional interpretation enables a better understanding of interpretation in practice. It explains why a mid-century court might insist it was following the plain meaning of the Constitution when it presumed the constitutionality of an expansive view of federal power. Or why a late-century court could have ignored the plain meaning rule and made ample use of extratextual sources that indicated a clause's purpose. Or why the *Lochner* Court reasoned that the Constitution protected the liberty of contract.

---

[275] The model may have been generally abandoned in new treatises after 1868, but it continued to appear in updated editions and reprints of the leading works of the school. *See, e.g.*, THOMAS COOLEY, A TREATISE ON THE CONSTITUTIONAL LIMITATIONS WHICH REST UPON THE LEGISLATIVE POWER OF THE STATES OF THE AMERICAN UNION (7th ed. 1903).

The most striking observation to be drawn from this history is that today's purposivism, far from being a twentieth-century oddity, follows an intentionalist and purposivist tradition dating at least to Blackstone and continuing throughout the nineteenth century.

The two purposivist approaches were the conservative choices, reflecting a far greater respect for both original intent and text than evolutionary constitutionalism. The purposivists looked beyond the text for indicia of intent, but, unlike their evolutionary successors, never suggested dispensing with it. Evolutionary constitutionalism, by contrast, pushed the outer bounds of interpretation and suggested ignoring text and intent altogether.

The deeply rooted intentionalism that dominated the nineteenth century appears to have precluded textualist constitutional interpretation through the majority of the century. If the treatises discussed here are any indication, there was in the nineteenth century little, if any, currency to the proposition that the Constitution is to be read with a narrow focus on its text.

This does not preclude the possibility that textualist constitutional interpretation is consistent with originalism. It is possible that jurists of the founding generation read federal and state constitutions with a textualist lens, only for the practice to fall out of fashion without being memorialized in a treatise. It appears that whether such a practice existed to any significant degree can be determined only through a systematic review of case law.

If there was no such trend, it appears that in constitutional interpretation, it is textualism and not purposivism that is a twentieth-century creation. This does not mean that on the merits, textualism is wrong or purposivism right. The historical predominance and relative conservatism of purposivism in the nineteenth century cannot determine our choices today.

It does, however, suggest that contemporary textualism is not a renaissance. It is not the default choice and appears not to be the conservative approach. Rather, it is a bold, novel departure from a well-established purposivist tradition. An awareness of this tradition is particularly relevant today as courts and scholars face the prospect of a textualist consensus.

THE PEACEMAKER TEST: APPLICATION AND COMPARISON

Luther T. Munford[*]

DESCRIPTION OF THE PEACEMAKER TEST 641
- *Compel and control participation* 641
- *Honor and respect the litigants* 641
- *Command the respect of the litigants and the public* 642
- *Give a final decision* 642
- *Promote reconciliation* 642

PAST APPLICATION 643
- *Recovery for the loss of the enjoyment of life caused by death* 643
- *Rule against vacating a judgment to facilitate settlement* 643
- *Rule against restrictions on judicial campaign speech* 644

FURTHER APPLICATION OF THE PEACEMAKER TEST 644
- *The problem of uncertainty: Suits for emotional disturbance* 644
  - Recovery for fear of future illness 646
    - *Peacemaker Analysis* 648
  - Caps on damages for pain, suffering, and emotional distress 651
    - *Peacemaker Analysis* 654
- *Sanctions for an "improper purpose" of seeking settlement "pressure"* 661
    - *Peacemaker Analysis* 665
- *Lawyer advertising* 670
    - *Peacemaker Analysis* 674

CONCLUSION 680

---

[*] Phelps Dunbar LLP, Jackson, Mississippi. The author is indebted to Virginia Munford, Virginia Kerr, and Gregg Mayer for their critical comments on drafts of this article, and to Gary Thompson and Tracey Ezell for their editorial help. Copyright Luther T. Munford 2010.

> Perhaps in time legal philosophers will cease to be preoccupied with building "conceptual models" to represent legal phenomena . . . and will turn instead to an analysis of the social processes that constitute the reality of law.[1]

The "Peacemaker Test" executes the "turn" that Lon L. Fuller, in the passage quoted above, hoped legal philosophy would take. Although the Peacemaker Test is what Fuller called a "conceptual model," it was derived from the work of peace studies scholars who described and listed the characteristics that they observed in effective peacemakers. Because of its origin in actual dispute resolution, it is anchored in what Fuller called the "social processes that constitute the reality of law."[2]

The test asks questions about legal rules in order to determine whether the rules foster effective and easy settlement of disputes that arise under the rules. A previous article developed the test, described legal doctrines that reflect the characteristics essential to peacemaking, and then assessed the peacemaking characteristics of three illustrative legal rules.[3] Those rules, when examined, were found to invite speculation, encourage contention, or engender disrespect for the courts, and so did not foster peacemaking.

This article briefly reintroduces the test, and then applies it to four other legal rules. The rules considered here are the rules that 1) allow recovery for a risk of future disease; 2) cap awards for emotional distress and other non-economic damages; 3) inflict sanctions on lawyers for actions taken with an "improper purpose;" and 4) permit lawyer advertising that uses both gimmicks irrelevant to the practice of law and self-laudatory mottos.

---

[1] LON L. FULLER, THE MORALITY OF THE LAW 242 (Yale Univ. Press rev. ed. 1969).

[2] *Id.* Fuller called the study of those processes "eunomics," or "the science, theory, or study of good order and workable social arrangements." Lon L. Fuller, *American Legal Philosophy at Mid-Century*, 6 J. LEGAL EDUC. 457, 477 (1954). *See also* THE PRINCIPLES OF SOCIAL ORDER: SELECTED ESSAYS OF LON L. FULLER (Kenneth I. Winston ed., rev. ed. 2001).

[3] Luther T. Munford, *The Peacemaker Test: Designing Legal Rights to Reduce Legal Warfare*, 12 HARV. NEGOT. L. REV. 377, 382 (2007).

DESCRIPTION OF THE PEACEMAKER TEST

The "Peacemaker Test" asks five broad questions about each legal rule:

1. Does it limit participation to the parties most interested in the quarrel?
2. Does it encourage respect for the parties?
3. Does it encourage party and public respect for the courts and their officers?
4. Does it promote the finality of the decision that ends the quarrel?
5. Does it help the parties reconcile and form a more co-operative relationship?

Each of these questions captures qualities that social scientists have listed in descriptions of good peacemakers.[4] Each of them also finds expression in existing legal doctrines.

Compel and control participation

A good peacemaker needs control over who participates in the peacemaking process. The peacemaker is a "convenor" and "facilitator" who can bring the necessary parties to the table and keep unnecessary external patrons away from it. Legal doctrines of third-party practice, intervention, joinder, severance, and standing give a court control over who participates.[5]

Honor and respect the litigants

A good peacemaker respects the litigants. The peacemaker helps them explore and discover ways the suit might be resolved. The peacemaker gives them the power to resolve the dispute and protects them from unnecessary litigation injury, i.e., the psychological as well as financial costs of a lawsuit. Under this heading fall the legal doctrines of due process, discovery, professional responsibility, and the litigants' right to be heard in court.[6]

---

[4] *Id.* at 390 & n.55.

[5] *Id.* at 391.

Command the respect of the litigants and the public

Respect for the peacemaker legitimizes the result of the dispute. Legal authority confers legitimacy. So does the method by which judges are selected. Justices come to the Supreme Court only after presidential nomination and Senate confirmation, and they enjoy life tenure. But custom and charisma also contribute to legitimacy. Justices wear black robes and hold court in a marble temple. They speak only through reasoned decisions whose logic earns respect.[7] Courts also gain legitimacy when they entertain the parties' views, act without bias, and stick to the evidence presented.[8]

Give a final decision

A successful peacemaker must be able to monitor and enforce its decision. The doctrines of claim and issue preclusion end litigation. Courts can hold those who disobey their specific orders in contempt. Even emotional and financial exhaustion reinforce finality and, in this respect, can play a positive role.[9]

Promote reconciliation

A reconciler identifies the interest and values at stake and helps the parties find an acceptable solution. A successful peacemaker does that and may even build new relationships between adversaries. Litigation does not promote reconciliation, but settlement may. To that end, courts may require mediation before trial. Sometimes settlement discussions can lead to a result that leaves both sides better off. Settlement also creates the possibility of apology.[10]

---

[6] *Id.* at 391-93.

[7] *Id.* at 393-95.

[8] TOM R. TYLER, WHY PEOPLE OBEY THE LAW 163-64 (1990).

[9] Munford, *supra* note 3 at 395.

[10] *Id.* at 395-96.

This test measures the ease with which parties to a dispute under a rule will be able to settle that dispute either without litigation or with litigation that inflicts minimal litigation injury. It aims to achieve the goal identified by David Wexler, i.e., to make "the law itself . . . preventative."[11] Wexler, a founder of the "therapeutic jurisprudence" movement, has expressed the hope that through "process-oriented legal doctrines" law could "reduce[] the likelihood of future legal conflict and resort to litigation."[12] That is the goal of the Peacemaker Test.

PAST APPLICATION

A previous article used the Peacemaker Test to examine three legal doctrines or rules.[13]

Recovery for the loss of the enjoyment of life caused by death

Traditional wrongful death damages compensate the heirs for their financial losses and loss of society. The additional ability to recover money for speculative post-death hedonic damages threatens defendants with catastrophic financial loss, telegraphs to the public the arbitrariness of the judicial system, and even without these effects, provides a monetary incentive to file speculative lawsuits.[14]

Rule against vacating a judgment to facilitate settlement

The rule is premised on a belief that judicial decisions have value to the public as well as to the parties. That may be true, but it is less likely to be true when the decision is so novel that one side wants to avoid having it stand as a future precedent. If the novel rule is worthy, it will be followed by other courts in the future whether or not it stands as a precedent in the court that originally adopted it. Vacating a judgment to encourage settlement thus would not unduly damage the interests of third parties, would honor the litigants, and would increase the possibility for reconciliation.[15]

---

[11] David B. Wexler, Introduction, *Therapeutic Jurisprudence in the Appellate Arena*, 24 SEATTLE U. L. REV. 217, 222 n.31 (2000) (emphasis omitted).

[12] *Id.*

[13] Munford, *supra* note 3.

[14] *Id.* at 397-401.

Rule against restrictions on judicial campaign speech

While the rule does not affect participation or finality, it encourages an appearance of partiality by allowing judicial candidates to take sides on controversial political issues. That diminishes the respect owed to the parties, undermines respect for the courts, eliminates one reminder to candidates of their duty to be impartial, and so weakens the court's ability to make peace among the parties.[16]

FURTHER APPLICATION OF THE PEACEMAKER TEST

The problem of uncertainty: Suits for emotional disturbance

Legal rules that allow lawsuits based on uncertain criteria empower the judge and jury but impede settlement. For that reason, they promote litigation that inflicts unnecessary "litigation injury," i.e., damage to reputation and mental well-being necessarily caused by adversarial litigation.[17] Because emotional disturbance knows no yardstick, courts have historically viewed suits for emotional disturbance with great skepticism.

Two examples, one specific and one general, can be used to test that skepticism. The specific example is the recovery of damages for fear of contracting a future disease. The general example is the attempt to cap emotional distress and other non-economic damages in a personal injury suit.

Looked at from a peacemaker point of view, damages for emotional disturbance bear some resemblance to post-death hedonic damages. Damages for emotional disturbance include damages for pain, suffering, and emotional distress. Those damages, like post-death hedonic damages, increase the risk that a judgment will be beyond the means of even a wellinsured defendant. The possibility of such a large judgment encourages filing of suits. At the same time, its unpredictable size impedes settlement.

---

[15] *Id.* at 402-06.

[16] *Id.* at 407-13.

[17] *Id.* at 382-88.

But emotional disturbance damages also differ from hedonic damages. Even the most extreme skeptic must admit that emotional disturbance damages do exist, especially when the person who sues has also suffered a significant physical injury. These damages do compensate a living person for a real injury. To that extent they properly deter wrongful conduct. So the peacemaker analysis of mental injury damages has to differ from the analysis of hedonic damages.

Suspicious of false claims, the courts at first refused stand-alone claims for emotional disturbance. A century or so ago, however, the courts began to allow a person who sued for "bodily harm" to also recover for emotional disturbance causally related to that harm or the underlying wrong. But the recovery was only for "physical harm resulting . . . from fright or shock or other similar and immediate emotional disturbance."[18] In other words, courts allowed recovery for anxiety that caused physical harm, but, absent that harm, allowed no recovery.[19]

In the early 1970s, the American Law Institute revised its restatement of personal injury law and expanded the number of situations in which a party could recover for emotional disturbance. It eliminated the requirement that the emotional disturbance itself must have caused "physical harm." It also dropped the immediacy requirement. So, for example, in disfigurement cases the new rule allowed recovery for humiliation, worry, and grief even if they did not cause illness and occurred long after the accident.[20]

---

[18] RESTATEMENT (FIRST) OF TORTS § 456 (1934).

[19] RESTATEMENT (FIRST) OF TORTS § 905 cmt. e, illus. 9 (1939). The illustration says that someone who used a knowing falsehood to inflict emotional distress was liable only if the distress caused physical harm.

[20] Norfolk & W. Ry. Co. v. Ayers, 538 U.S. 135, 149 (2003); RESTATEMENT (SECOND) OF TORTS § 456 cmt. d (1965).

Recovery for fear of future illness

The U.S. Supreme Court has rejected stand-alone emotional distress claims brought by workers who were exposed to asbestos but who, as yet, had no asbestos-related diseases. It did so in part for peacemaker reasons. It said that uncabined recognition of claims for negligently inflicted emotional distress would hold out the "very real possibility of nearly infinite and unpredictable liability" for those sued.[21]

Against that backdrop, the court in *Norfolk & Western Railway Co. v. Ayers*[22] undertook to decide whether workers exposed to asbestos could recover for fear of an increased risk of getting cancer if they suffered from some other asbestos-related disease, such as asbestosis.[23] The six workers, who sued under a federal liability statute, were all more than 60 years old. Five of them had a long history of cigarette smoking.[24] People who do not smoke run a 22% chance of dying of cancer.[25] If a member of that same group is diagnosed with asbestosis, the chance of dying of cancer rises to about 33%.[26] There is, however, no medical connection. Asbestosis does not lead to cancer.[27] The "increased risk" is just a statistical association.

Moreover, the diagnosis of asbestosis is freighted with uncertainty. Signs of asbestosis on a chest x-ray resemble more than 100 other lung disorders.[28] The parts of the diagnostic test not dependent on those signs test breathing capacity, which the worker can manipulate and which may relate to other diseases, such as those caused by tobacco.[29] In *Ayers*, the expert for the six workers diagnosed asbestosis from nothing but exposure history.[30]

---

[21] Consol. Rail Corp. v. Gottshall, 512 U.S. 532, 546 (1994).

[22] 538 U.S. 135 (2003).

[23] *Id.* at 140.

[24] *Id.* at 167 (Kennedy, J., concurring in part and dissenting in part).

[25] *Id.* at 184 (Breyer, J., concurring in part and dissenting in part).

[26] *Id.*

[27] *Id.* at 153-54 (majority opinion); *see also id.* at 172-73 (Kennedy, J., concurring in part and dissenting in part).

[28] Brief of Petitioner at 4, Norfolk & W. Ry. Co. v. Ayers, 538 U.S. 135 (2003) (No. 01-963).

[29] As a result, one 1990 study found that only about 3.6% of the asbestosis diagnoses it studied were correct. R. B. Reger, et al. *Cases of Alleged Asbestos-Related Disease: A Radiologic Re-Evaluation*, 32 J. OCCUPATIONAL MED. 1088, 1089 (1990).

The workers' only claims in the suit were that they suffered shortness of breath and feared that they would get cancer in the future. With respect to fear—their sole emotional disturbance claim—five said they had a "concern" they might get cancer in the future, or worried about it, while one said nothing at all.[31] The jury nonetheless awarded the workers from $770,000 to $1.2 million each, which the trial judge reduced somewhat.[32]

The Supreme Court ultimately held that it did not matter if the two diseases were not causally related. All that mattered was that the fear of developing cancer arose from the same underlying act, i.e., past contact with asbestos.[33] Because only 18% of those exposed to asbestos contracted asbestosis, the potential for liability was much smaller than it would have been without a rule that only those with asbestosis could sue.[34] The Court majority did not share the concern of the dissenting Justices who pointed out that asbestos litigation has already driven 57 companies into bankruptcy and that large recoveries for slightly-injured workers could mean there would be no money left to compensate seriously injured workers in the future.[35]

The Supreme Court also refused to require that those who claimed fear of future cancer show some physical manifestation of their fear. The Court said fear claims could go forward if the jury was convinced that the fear was "genuine and serious," even though no objective evidence was offered to support the fear claim, and even though the jury did not have to find that the increased risk of cancer actually existed.[36]

---

[30] Joint Appendix at 198-99, 229, Norfolk & W. Ry. Co. v. Ayers, 538 U.S. 135 (2003) (No. 01-963) [hereinafter J.A.].

[31] Brief of Petitioner, *supra* note 28, at 9.

[32] J.A., *supra* note 30, at 588, 590-613.

[33] Norfolk & W. Ry. Co. v. Ayers, 538 U.S. 135, 153-55 (2003).

[34] *Id.* at 157 (citing W. KEITH, C. MORGAN & ANTHONY SEATON, OCCUPATIONAL LUNG DISEASES 319 (3d ed. 1995)).

[35] *Id.* at 168-69 (Kennedy, J., concurring in part and dissenting in part).

[36] *Id.* at 157 (majority opinion). See also CSX Transp. Inc. v. Hensley, 129 S. Ct. 2139 (2009), in which the court reversed a jury verdict because the instructions did not require the jury to find that the fear was "genuine and serious." *Id.* at 2141.

This drove the dissenters to eloquence. Justice Anthony Kennedy said: "If instructing a jury to calculate an increased risk of cancer invites speculation, then asking the jury to infer from its estimate a rough sense of fear based on the risk invites speculation compounded."[37] Justice Stephen Breyer noted the irony in a claim that a risk-taking smoker, who by smoking has doubled his chance of cancer death, would be afraid of a marginal increase in that risk. The real problem, he believed, was the "impossibility of knowing an appropriate compensation for asbestosis insofar as its appearance tears away that veil of disregard that ordinarily shelters most of us from fear of cancer, if not fear of death itself."[38]

Peacemaker Analysis

Compel and encourage participation. If claims for fear of contracting a disease in the future are phony, then rewarding those claims with damages brings into court people who should not be there. To that extent, these damages ignite disputes rather than settle them and should not be allowed. The question becomes the ability of the courts to separate true fear claims from false fear claims.

Certainly the traditional English and early American court systems did not have much confidence in the courts' ability to detect the truth when the stakes were high and the proof depended entirely on oral testimony. Evidence of this is found in the statute of frauds, which denies enforcement to certain important transactions, such as the sale of land, if they are not in writing.[39] Other examples include the requirement that clear and convincing evidence must support a fraud claim;[40] the rule that bars oral testimony that is contrary to the words of a written contract;[41] the statute that banned certain testimony about transactions with people now dead;[42] and, even the Old Testament rule that a sworn complaint could be overcome only by the testimony of two witnesses.[43]

---

[37] *Norfolk & W. Ry. Co.*, 538 U.S. at 179 (Kennedy, J., concurring in part and dissenting in part).

[38] *Id.* at 185 (Breyer, J., concurring in part and dissenting in part).

[39] RESTATEMENT (SECOND) OF CONTRACTS § 110 (1981) (listing classes of contracts subject to writing requirement).

[40] 32A C.J.S. *Evidence* § 1624 (2008) (stating that a heightened proof standard is used "when fundamental rights are involved and the legal and social ramifications of the civil proceeding are serious").

Today, some of these rules no longer exist or are no longer enforced because of the occasional injustice they can cause. But there is little reason to believe persons who sue or are sued today lie less frequently. Nor is there any reason to believe that courts today are any better than predecessor courts in discerning the truth. Even discovery, the celebrated modern engine of truth-finding, may do little good when discovery of relevant facts is impossible because those facts lie—whether a person is afraid or not—inside an opposing party's head.

In fact, not only is it difficult to ascertain whether a fear of cancer is "genuine and serious," but also the ability to sue for fear of developing cancer is likely to greatly increase the chances of success on the merits and so expand litigation whether the claim has merit or not. Without a fear of cancer claim, a suit for shortness of breath may have little appeal to a jury. Bringing cancer into the case, however, introduces a dreaded disease and invites testimony about its catastrophic horrors. That testimony appeals to the jury's emotions in both rational and irrational ways. Ironically, the more a defense expert testifies about how low a risk is, the more the jury is likely to believe the risk is high.[44] But for the claim for fear of the risk, all of this would be irrelevant and inadmissible.

Respect for the litigants. Recoveries for fear honor the injured person who claims them, but at a considerable price paid in litigation injury. Whether they dishonor the person who has caused the injury is a question of amount, which is addressed below.[45]

---

[41] RESTATEMENT (SECOND) OF CONTRACTS § 213 cmt. b, illus. 1 (1981) (binding written agreement supersedes prior oral agreement).

[42] 98 C.J.S. *Witnesses* § 215 (2002) (dead man's statutes).

[43] *See* Munford, *supra* note 3, at 392 n.69 (citing *Deuteronomy* 19:15 & *Matthew* 18:15-20).

[44] W. Kip Viscusi, *Jurors, Judges, and the Mistreatment of Risk by the Courts*, 30 J. LEGAL STUD. 107 (2001).

[45] *See infra* pp. 17-21.

Command respect from the litigants and the public. The Supreme Court's willingness to allow a recovery for fear arising from a slight statistical increase in the risk of death encourages disrespect for the courts. Justice Breyer, one of the first scholars to consider the effect of modern risk evaluation studies on the law,[46] is right. People operate with a "veil of disregard" for future risks. It is not necessarily correct to say that a person who has smoked tobacco for decades cares very much about a much smaller increased risk of death from cancer. It is even more unlikely that a court could measure such an increment.

Moreover, recovery for fear of an increased cancer risk is a tort remedy that only benefits those who are afraid of something that is not going to happen to them.[47] Under the "second disease rule," workers who in fact get cancer can sue at that time and recover for the medical and emotional consequences of their illness.[48] They do not need a recovery for fear. Only those who will never get cancer benefit from the recovery for fear. This is yet another reason the remedy may engender public disrespect.

Finality. The Supreme Court's decision in *Ayers* does not by itself affect the finality of lawsuits that are filed.

But it is noteworthy that the Court did not cut off a later recovery in a separate suit for cancer by those who eventually got cancer. Accordingly, the availability of a recovery for fear does not prevent relitigation of the issues in a "separate disease" suit. The first recovery may diminish the appetite for the second, but that would only be true where the cancer damages were small or the causation was doubtful.

Reconciliation. Emotional distress damages awarded in the absence of objective proof make little sense to the person who is sued. They drive the parties farther apart rather than closer together. The availability of those damages may also lead the person who sues to have unrealistic expectations about how much that person can recover. Perceived unfairness is a principal source of litigant dissatisfaction with the court system—even more than cost and delay.[49]

---

[46] STEPHEN BREYER, BREAKING THE VICIOUS CIRCLE: TOWARD EFFECTIVE RISK REGULATION (1993).

[47] *See* Norfolk & W. Ry. Co. v. Ayers, 538 U.S. 135, 153 (2003) ("gap" in liability would exist if those who feared cancer but did not get it were not afforded compensation).

[48] Pursuant to the "second disease rule" courts have allowed a second case to be brought by the worker who later develops cancer. *See id.* at 152-53; *see also, e.g.*, Marinari v. Asbestos Corp., 612 A.2d 1021, 1022-23, 1028 (Pa. Super. Ct. 1992); Sopha v. Owens-Corning Fiberglas Corp., 601 N.W.2d 627, 642 (Wis. 1999).

From a peacemaker point of view, *Ayers* creates a cause of action that allows potentially ruinous recoveries for injuries whose existence cannot be ascertained without speculation or caprice. The cause of action for fear encourages exaggeration or even perjury. It also falsely assumes a degree of perfection in the truth-telling abilities of the courts that does not exist in the real world. It encourages suits for fear while leaving open the possibility of a second suit if cancer actually develops. None of that promotes peacemaking. A rule that allowed damages for emotional disturbance only when cancer actually developed would avoid many of these problems.

Caps on damages for pain, suffering, and emotional distress

If a person's carelessness has caused a physical injury accompanied by pain, suffering, and emotional distress, the task for the jury is to put a dollar amount on that injury. This raises still more questions when examined from the peacemaker point of view. For the purpose of this section, pain, suffering, and emotional distress, none of which are susceptible to precise economic measurement, will be called "non-economic" damages.

These damages do at least two things:

First, they compensate the injured person for that person's emotional disturbance injuries, including pain and suffering.

Second, they award the injured person extra money so that person can pay a lawyer's fee without having to sacrifice money needed for medical bills or lost pay. So, for example, a person with $50,000 in medical bills and $100,000 in lost income might recover three times the medical bills, or $150,000 for non-economic damages. These three numbers added together would total $300,000, from which the attorney might take 40%, or $120,000, leaving the person injured with a $180,000 recovery, which is enough to cover the medical bills and lost income, and then some.

---

[49] E. ALLEN LIND, ET AL., THE PERCEPTION OF JUSTICE: TORT LITIGANTS' VIEWS OF TRIAL, COURT-ANNEXED ARBITRATION, AND JUDICIAL SETTLEMENT CONFERENCES 75-80 (1989), http://www.rand.org/pubs/reports/R3708/ (follow "Full Document" hyperlink).

Courts seldom, if ever, explicitly acknowledge this second purpose. Our legal system supposedly follows the "American Rule," which requires each party to a case to pay the party's own lawyers' fees. That rule ensures that neither litigant can expect to make a profit out of litigation. But both insurance adjusters and legal commentators have had no trouble linking the recovery of non-economic damages to lawyer fees.[50] The logic is simple. The courts that once considered contingent fees unethical have allowed them so that poor people can sue for their injuries. It would frustrate that purpose, to a degree, to force the poor person to give up a large fee out of what was needed to make that person financially whole.

Traditional limits. The established rules of appellate review require an appellate court to affirm the amount of a jury award unless the amount is so high as to shock the judicial conscience or is "[d]isproportionate to the injury sustained."[51] The effect of these gentle admonitions depends entirely on the views of the judges who enforce them. At times, judges working under this rubric have limited non-economic damages either by using the insurance adjuster's rule that in personal injury cases those damages should be from three to five times the amount of the injured's medical bills,[52] or by a comparison to what courts in the state or in other states have done in similar cases, or by simply relying on their intuition to decide how much is too much.[53]

---

[50] *See* DAN B. DOBBS, LAW OF REMEDIES: DAMAGES-EQUITY-RESTITUTION § 3.1, at 211 (2d ed. 1993) (citing attempt to cover attorney's fees as "the most important single reason why non-pecuniary losses support such large dollar recoveries"); Stanley Ingber, *Rethinking Intangible Injuries: A Focus on Remedy*, 73 CAL. L. REV. 772, 811 (1985); Jeffrey O'Connell, *A Proposal to Abolish Payment for Pain and Suffering in Return for Payment of Attorney's Fees*, 1981 U. ILL. L. REV. 333, 351 (1981) (settlement of three times the medical bills yields one-third for the lawyer, one-third for the doctor, and one-third for the client).

[51] 19 JAMES WILLIAM MOORE ET AL., MOORE'S FEDERAL PRACTICE ¶ 206.08 (3d ed. 2009).

[52] H. LAURENCE ROSS, SETTLED OUT OF COURT: THE SOCIAL PROCESS OF INSURANCE CLAIMS ADJUSTMENTS 108 (1970); O'Connell, *supra* note 50, at 334 (insurers pay three times medical bills).

Damage caps. Businesses, charities, and insurance companies, however, have reacted to the growth of non-economic damages by encouraging state legislatures to adopt statutes that cap them at a certain amount, for example, at $250,000. At least 29 states have adopted one type of cap or another.[54] In state courts, however, the question has arisen as to whether these caps are constitutional under various provisions of each state's constitution.[55]

For example, the Oregon legislature capped non-economic damages at $500,000 for each person injured. In the case that tested that cap, *Lakin v. Senco Products, Inc.*,[56] a pneumatic nail gun fired accidentally, shot a nail into a worker's brain, and crippled him. It paralyzed his left arm and leg and changed his personality. He could no longer think clearly or express emotions. He could no longer live independently. The jury ruled that the nail gun manufacturer should pay punitive damages in addition to $3.3 million for his economic damages, which included medical bills and lost wages. He also received $2 million for his non-economic damages, and his wife received $876,000 for her emotional damages. The trial judge honored the statute and reduced the non-economic damages to $500,000 for the husband and $500,000 for the wife, for a total of $4.3 million, plus punitive damages.[57]

---

[53] *See* Gasperini v. Ctr. for Humanities, Inc., 518 U.S. 415, 425 (1996) (New York requires comparison to awards approved in similar cases); Moore v. Angela MV, 353 F.3d 376, 385 (5th Cir. 2003) (applying Fifth Circuit's "maximum recovery rule" which limits awards to within 133% of comparable damage awards). Such an approach has been said to lead to a "progressive maximization" of damages. Louis L. Jaffe, *Damages for Personal Injury: The Impact of Insurance*, 18 LAW & CONTEMP. PROBS. 219, 234 (1953).

[54] *See* Catherine M. Sharkey, *Unintended Consequences of Medical Malpractice Damages Caps*, 80 N.Y.U. L. REV. 391, 496-97 (2005) (listing 29 states having either caps on medical malpractice damages or general caps that would apply in medical malpractice cases).

[55] *See generally* Matthew W. Light, Note, *Who's the Boss?: Statutory Damage Caps, Courts, and the State Constitutional Law*, 58 WASH. & LEE L. REV. 315 (2001).

[56] 987 P.2d 463, 466 (Or. 1999).

The Oregon Supreme Court unanimously found the cap statute unconstitutional and reinstated the original jury award. It reasoned that the state constitution protected the right to trial by jury and, in 1857 when the state adopted the constitution, juries—not judges—determined the amount of damages in civil cases. From that, the court found that the legislature could not take away from a jury the power to fix the amount of damages in a civil case.[58] Finally, it relied on decisions of courts in other states that had declared similar caps unconstitutional.[59]

The decision thus affirmed an award that gave a $6.1 million recovery, not considering punitive damages. That gave them more than enough to pay their attorney a 40% contingent fee, or $2.4 million, keep their $3.3 million in economic damages, and still have $400,000 left over.

Peacemaker Analysis

Participation. When a court awards large amounts of money for pain and suffering and other emotional damages, it encourages lawsuits and brings more parties into court, particularly if parties cannot predict in advance what a verdict might be.

Conversely, a cap discourages suits. The most frequent argument against caps comes from lawyers who say they will not be able to take some cases if caps limit the recovery. If there was any doubt about the relationship between non-economic damages and attorneys' fees, that argument eliminates it.

But the discouragement is not quite as much as some might think. The cap does not restrict awards for economic damages, including present and future medical expenses, or punitive damages. In other words, if emotional distress has created a future need for medicine or psychological help, the estimated fees for those services will not count against the cap.[60]

---

[57] *Id.* at 466. The punitive damages award was $4 million. It is ignored here to simplify the description.

[58] *Id.* at 472-73. The court admitted that the state constitution allowed the legislature to abolish causes of action, but found that irrelevant. *Id.* at 473-74. The court did not consider whether Oregon law allowed the recovery of non-economic damages in 1857 in a tort case.

[59] *Id.* at 474-75.

If logic prevails, clients and lawyers will sue as long as their estimated cost of recovery sufficiently exceeds the cost of bringing suit. For the client, any money exceeding expenses and compensation for expected litigation injury will make the client better off. Setting those items aside for simplicity,[61] the question is how high the total award would have to be before some percentage of attorneys' fees, say 40%, would be greater than the $500,000 non-economic damage cap. If $.40 \times X = \$500{,}000$, then $X = \$500{,}000 / .40 = \$1{,}250{,}000$. As a result, 40% of the total award is only greater than the $500,000 cap when the total award exceeds $1.25 million. The $500,000 cap will not keep a plaintiff from having enough to pay a 40% attorney's fee out of non-economic damages unless the total award for both economic and capped non-economic damages exceeds $1.25 million.

Those who favor limits on the recovery of non-economic damages have frequently recommended that the limits be balanced by giving plaintiffs' counsel the right to recover attorneys' fees in addition to recoverable damages.[62] That would ensure that limits on non-economic damages would not deprive a worthy tort victim of the victim's day in court. The American

---

[60] *See* Martin V. Totaro, Note, *Modernizing the Critique of Per Diem Pain and Suffering Damages*, 92 VA. L. REV. 289, 290, 304-20 (2006) (describing "revolution in pain management therapy over the past thirty years").

[61] For the same reason, this example also does not discuss health insurance, which usually carries with it an insurer's right to recoup its payments from the damage award.

[62] *See, e.g.*, II AMERICAN LAW INSTITUTE, ENTERPRISE RESPONSIBILITY FOR PERSONAL INJURY: APPROACHES TO LEGAL AND INSTITUTIONAL CHANGE 199-230 (1991) [hereinafter ENTERPRISE RESPONSIBILITY] (schedule damages and award fee); Ingber, *supra* note 50, at 812 (court-award attorneys' fees may be necessary to guarantee due process); O'Connell, *supra* note 50, at 369 (proposing to abolish non-economic damages but award attorneys' fees); Thomas D. Rowe, Jr., *Indemnity or Compensation? The Contract with America, Loser-Pays Attorney Fee Shifting, and a One-Way Alternative*, 37 WASHBURN L.J. 317, 319-20 (1998).

Law Institute has proposed that the fee be a percentage recovery contingent on the outcome of the case.[63] Using a percentage rather than just a "reasonable" fee would avoid controversy over reasonableness which, in other contexts, has been labeled "Frankenstein's monster" litigation.[64] For that reason, it would advance peacemaking.

To continue with the example above, suppose noneconomic damages were reduced to a $500,000 cap, leaving the person who sued with a damage recovery of $750,000 in economic damages and $500,000 in noneconomic damages, totaling $1.25 million. If a 66% fee were added to that total, then the person who sued would recover another $825,000 for a total of $2.075 million. The figure of 66% was chosen because, when it is used, the fee is always 40% of the total of the damage recovery plus the fee, i.e., $825,000 is 40% of $2.075 million. A higher ratio might be allowed when damages are small, but the right to sue is especially important. A lower ratio might be allowed when the economic damages are, as in *Lakin*, especially large. In other words, fee percentages could be scheduled.

Those who are sued would think this was unfair because they could not get their fees paid if they win. But that is, in effect, the way the system currently works. The payment of fees by those who are sued is also permitted by more than 100 federal statutes, which allow plaintiffs to recover attorneys' fees on terms more generous than those afforded defendants.[65] And in ordinary tort cases, those who are sued in effect pay "fees" in the form of non-economic damages that have no predictable cap. With a cap and a calculated fee, defendants would gain predictability and a greater possibility of settlement.[66] Whether they would lose money in an individual case would depend on the magnitude of the non-economic damage cap, its effect on the judgment, and the fee schedule.

---

[63] ENTERPRISE RESPONSIBILITY, *supra* note 62, at 297.

[64] Hensley v. Eckerhart, 461 U.S. 424, 455 (1983) (Brennan, J., concurring in part and dissenting in part).

[65] Harold J. Krent, *Explaining One-Way Fee Shifting*, 79 VA. L. REV. 2039, 2041 (1993).

[66] *See* Steven Shavell, *Suit, Settlement, and Trial: A Theoretical Analysis Under Alternative Methods for the Allocation of Legal Costs*, 11 J. LEGAL STUD. 55, 68 (1982) (under conditions of uncertainty, one-way fee shifting is more likely to produce settlement than no shifting at all).

Respect for the litigants. Once again, paying an injured person shows respect for that person, but the effect on that person would be far different if the person were a defendant. It is important to keep in mind that no one knows in advance on which side of a lawsuit he may find himself.

The part of the *Lakin* judgment discussed above did not turn on the net worth of the manufacturer defendant. Under the principles of that case, the same sort of damages could be awarded against any defendant who caused a serious brain injury, be it an automobile driver, a small independent contractor hired for a specific job on a construction site, or the hardware store that accidentally lost safety instructions that were supposed to go with a power tool. Even if that store owner had a $5 million umbrella insurance policy, the judgment could still bankrupt the owner.

On the tort victim's side, it is worth asking whether victims in fact are already compensated for non-economic injuries when they recover for the cost of medical treatment for those injuries. Current law on non-economic damages dates back to a time before the invention of medicines now available to ameliorate mental injuries. Today, health professionals and medicine treat those injuries.[67] So conditions that previously would justify non-economic damages now support compensation for economic damages. As a result, there may be less need for the recovery of non-economic damages for either compensation or deterrence.

Awarding only the costs of medical treatment for non-economic injury could also lessen the degree of emotional injury. It would encourage victims to seek medical diagnosis and treatment, and so discourage the current temptation to validate legal claims by remaining untreated until the lawsuit is over.[68]

---

[67] *See* Totaro, *supra* note 60, at 290, 304-20 (discussing pain analysis and management).

Courts in the past have altered legal standards when advances in medicine changed the assumptions on which those standards rested. Examples include the standards for witness competency[69] and even for the definition of death.[70] Those alterations provide precedent for a change in the standards of emotional distress recovery to take into account increases in economic damages caused by better techniques for treating mental problems.

Command respect from the parties and the public. Critics of non-economic damages focus their aim on the courts' failure to use a predictable standard when computing the amount of these damages. The public wants like cases to be treated alike, and the subjective nature of non-economic damages has, for better or worse, brought them into disrepute.

For example, the American Law Institute has noted that non-economic damages compose roughly 50% of all medical malpractice and product liability awards, but their amounts are "erratic and unpredictable."[71] A leading scholar of American tort law has called them the "most uncertain element" in computing tort damages and, in his own review of Illinois awards, has found no correlation between the amount awarded and the nature of the injuries.[72]

Because non-economic damages are subjective, forensic techniques can easily affect the amount awarded. In the absence of objective numbers, arguments that a certain amount of money should be awarded for each day of future pain and suffering anchor numbers in the jury's mind and so artificially influence the verdict.[73] One study found that what psychologists call the "endowment effect"—the tendency to value losses more than equivalent gains—could double non-economic damages awarded by mock juries. If jurors were asked how much they would have to be paid to give up good health and suffer like the plaintiff, they would award twice as much as jurors who were only asked how much the plaintiff should be paid in the future for the plaintiff's existing injuries.[74] A foot injury forces a patient to lie in bed and think about nothing but the pain in the foot, and, as W. H. Auden put it, ". . . [W]ho when healthy, would become a foot?"[75]

---

[68] Bruce J. Winick, *The Jurisprudence of Therapeutic Jurisprudence*, 3 PSYCHOL. PUB. POL'Y & L. 184, 202 (1997) (discussing problem of "incompetency labels" and seeking alternatives).

[69] State v. Garcia, 658 A.2d 947, 966-67 (Conn. 1995).

[70] *See* State v. Guess, 715 A.2d 643, 649 (Conn. 1998); Narbeh Bagdasarian, *A Prescription for Mental Distress: The Principles of Psychosomatic Medicine with the Physical Manifestation Requirement in N.I.E.D. Cases*, 26 AM. J.L. & MED. 401, 431-33, 432 n.348 (2000) (discussing the "justification[s] for allowing scientific and medical advances to change and modify legal tests").

[71] *See* ENTERPRISE RESPONSIBILITY, *supra* note 62, at 201.

[72] O'Connell, *supra* note 50, at 342, 362.

It is also true that a given level of physical pain will produce different levels of emotional distress in different people based on factors that have nothing to do with the justice of the case. People with different personality types express the same pain differently. A neurotic will make himself miserable. A person with a character disorder will make everyone else miserable.

In addition, an effective lawyer can still argue to the jury that a person who shows no emotional injury on the witness stand has nevertheless suffered such an injury. The lawyer can tell the jury: "You can see what kind of a person my client is. He's not a whiner. He puts his chin up and holds it up. But you saw the lines on his face. Deep inside he hurts in a way we cannot imagine." The lawyer may be right. The lawyer may be wrong. It is hard to see how a jury could tell the difference. Just such an argument was made by experts who testified for the injured workers in the *Ayers* case.[76]

---

[73] Joseph H. King Jr., *Counting Angels and Weighing Anchors: Per Diem Arguments for Noneconomic Personal Injury Tort Damages*, 71 TENN. L. REV. 1 (2003).

[74] Edward J. McCaffery, Daniel J. Kahneman & Matthew L. Spitzer, *Framing the Jury: Cognitive Perspectives on Pain and Suffering Awards*, 81 VA. L. REV. 1341, 1351-52, 1372 (1995).

[75] W.H. AUDEN, *Sonnet from China XIV*, *in* COLLECTED POEMS 149, 154 (Edward Mendelson ed., Random House 1976).

[76] Brief of Respondents at 7, Norfolk & W. Ry Co v. Ayers, 538 U.S. 135 (2003) (No. 01-963). *See also* Totaro, *supra* note 60, at 297 (noting argument that a client's minimizing pain may indicate severe trauma).

If the jury can award emotional distress damages and be affirmed no matter how the injured person testifies, then the award of those damages is so capricious as to create even greater inequities among those who suffer such injuries. Public inequity strengthens public disrespect for the courts because lawsuits then appear to be more of a lottery than a wise resolution of a dispute. The uncertain nature of the recoveries also makes it more difficult to settle.[77]

For reasons like these, state legislatures have adopted damage caps that curb non-economic damages. But the caps carry with them their own form of arbitrariness. To a degree they simply substitute one form of arbitrariness for another. Instead of arbitrary damage sums they offer a single damage sum that will fit the facts of individual cases in arbitrary ways. They create the image of equal treatment without the reality. That problem would be mitigated by proposals, not yet adopted by any state, that would schedule non-economic damages according to the severity of the physical injury, the body part affected, and the age of the injured person.[78] In other words, to continue with the example above, the $500,000 cap would go up or down depending on those factors. That in turn would affect the amount of the attorneys' fees awarded. While not yet a part of tort law, schedules are commonly used to compute workers' compensation recoveries.[79]

Reconciliation. Damages that cannot be readily calculated provide room in negotiations for concessions that may save a settlement while all parties save face. However, without a yardstick, uncertainty creates disagreement, which usually makes it more difficult to settle because both sides are more likely to make mistakes when calculating the probable outcome of the case.[80]

---

[77] Linda Babcock & Greg Pogarsky, *Damage Caps and Settlement: A Behavioral Approach*, 28 J. LEGAL STUD. 341, 352, 362-63 (1999) (discussing the effect of caps, and the consequent increased in certainty, on the probability of settlement); O'Connell, *supra* note 50, at 352-58.

[78] *See generally* ENTERPRISE RESPONSIBILITY, *supra* note 62; Randall R. Bovbjerg, Frank A. Sloan & James F. Blumstein, *Valuing Life and Limb in Tort: Scheduling "Pain and Suffering*," 83 NW. U. L. REV. 908, 939 (1989).

[79] Bovbjerg, *supra* note 78, at 928.

From the peacemaker point of view, uncertainty over the amount that should be awarded to compensate a tort victim for non-economic damages makes it more difficult for the parties to settle, and so impedes peacemaking. What makes more sense from a peacemaking point of view would be to allow the recovery of medical treatment costs, scheduled non-economic damages, if any, and attorneys' fees calculated as a percentage of the total award. That would enable the parties to determine potential liability more accurately and so would aid settlement while properly respecting the interests of both the person who brings the suit and the person who is sued.

Sanctions for an "improper purpose" of seeking settlement "pressure"

Law designed to make peace helps minimize litigation injury. Litigation causes injury not only because it costs money, but also because the steps the adversary system requires often can damage the reputation, privacy, and emotional well-being of litigants and witnesses.[81]

The law generally protects lawyers from litigation injury. By necessity, a litigation privilege shields lawyers from the litigation injury they would suffer if the opposing party could sue them for those damages. The protection is particularly important because a lawyer not only can, but in some circumstances must, engage in activity that would otherwise be tortious. Lawyers file pleadings that injure the reputations of opposing parties. They use discovery to learn private facts. They cross-examine lying witnesses vigorously to inflict emotional distress that will bring out the truth.

The law grants lawyers immunity from suit for these acts because the adversary system requires lawyers to do these things. To take away immunity would make it impossible for anyone to afford to be a lawyer. Absent the privilege, after the entry of final judgment the winner's counsel could sue the loser's lawyer for tortious injury. Then the judgment in that suit could give rise to yet another suit of the same nature, and so on indefinitely. Litigation privilege prevents the follow-on litigation. The immunity also keeps the attorney from having a personal stake in the litigation. That frees the attorney to act solely in the client's interest while not sharing the client's emotional attachment to the suit. Because of the privilege, the attorney can play the dispassionate "decoupler" role peacemaking requires.[82]

---

[80] *See* Munford, *supra* note 3, at 402 & n.117.

[81] *See id.* at 382.

The principal exception to litigation privilege is the ability a party has to seek sanctions against opposing counsel under Rule 11 of the Federal Rules of Civil Procedure. That rule allows a court to sanction an opposing party or counsel if the party's legal argument is frivolous, if no evidence can be found to support a party's factual claims, or if a pleading the party files was motivated by an "improper purpose, such as to harass, cause unnecessary delay, or needlessly increase the cost of litigation."[83] Even a successful defense of such a claim undercuts litigation immunity, because the accused attorney must defend himself and so become, in effect, a party to the proceeding. Sanctions motions are infrequent, but when filed, they frequently target the opposing attorney rather than the opposing party. From 1993 to 2003, 75% of all federal sanctions awards were issued against attorneys.[84]

Rule 11 poses a particular threat to peacemaking because of its vagueness. Its vagueness creates an irony. Rule 11 ostensibly deters harmful litigation and prevents litigation injury. But the vagueness of the rule creates a potential for satellite litigation that increases the very litigation injury the rule is designed to prevent. It also inflicts that injury on opposing counsel rather than the opposing party.

---

[82] *See id.* at 391.

[83] FED. R. CIV. P. 11(b)(1).

[84] Peter A. Joy, *The Relationship Between Civil Rule 11 and Lawyer Discipline: An Empirical Analysis Suggesting Institutional Choices in the Regulation of Lawyers*, 37 LOY. L.A. L. REV. 765, 790 tbl.2 (2004).

For example, the requirement that a pleading not be so lacking in legal merit as to be "frivolous" would seem to be clear. In one critique, however, the author found four cases in which a federal district court had sanctioned an attorney for making a "frivolous" argument, yet within 12 months in other cases the U.S. Supreme Court found the same argument not only meritorious, but correct.[85]

The Rule 11 requirement that pleadings not have an "improper purpose" is even vaguer than the frivolity standard. In another context, ordinances penalizing an "improper purpose" have been found unconstitutional because a person of ordinary knowledge could only guess at the phrase's meaning.[86] The references to harassment and increased cost in the rule could provide some color to the meaning of improper purpose, but the courts have not confined it to those elements. In fact, the rule against acting with an improper purpose has been labeled the "vexatiousness" rule with the suggestion that anything designed to cause the opposing party distress, mental suffering, or annoyance might be enough.[87]

One line of "improper purpose" cases has even found sanctionable a motive that would normally be thought to promote peacemaking: a motive to take lawful actions that would "pressure" an opposing party to settle a case.[88] In *Federal Deposit Insurance Corp. v. Maxxam*,[89] the court upheld sanctions against the FDIC for pursuing a non-frivolous claim against the owner of a failed financial institution. The court said that at least one reason the FDIC pursued the claim was a desire to "pressure" the defendant to settle by conveying California redwood property he owned to the federal government.[90] The court held that even if the claim was properly filed, it was improperly pursued. In support, it quoted an FDIC lawyer's comment that it would be to the government's advantage to adopt a "moderately aggressive discovery plan of our own that inconveniences him and strengthens our case."[91] The court held that, even if the FDIC lawyers had mixed motives, some of which were proper, the presence of one improper motive was sanctionable if it caused the conduct.[92]

---

[85] Mark S. Stein, *Rule 11 in the Real World: How the Dynamics of Litigation Defeat the Purpose of Imposing Attorney Fee Sanctions for the Assertion of Frivolous Legal Arguments*, 132 F.R.D. 309, 318-20 (1990); *see also In re* McCabe, 583 N.E.2d 233 (Mass. 1991) (exonerating attorney sanctioned in NASCO, Inc., v. Calcasieu Television & Radio, Inc., 894 F.2d 696 (5th Cir. 1990), *aff'd on other grounds sub nom.* Chambers v. NASCO, Inc., 501 U.S. 32 (1991)).

[86] *See, e.g.*, Elam v. Bolling, 53 F. Supp. 2d 854, 862 (W.D. Va. 1999) (town ordinance prohibiting "improper conduct" unconstitutionally vague); Marin v. Univ. of P.R., 377 F. Supp. 613, 627 (D.P.R. 1974) (rule prohibiting "improper or disrespectful conduct" void for vagueness); Armstrong v. Ellington, 312 F. Supp. 1119, 1126 (W.D. Tenn. 1970) (statute prohibiting use of "other improper means" unconstitutionally vague); *see also* Gentile v. State Bar of Nev., 501 U.S. 1030, 1034, 1044-48 (1991) (legal ethics rule unconstitutionally vague).

[87] Ratliff v. Stewart, 508 F.3d 225, 235 n.13 (5th Cir. 2007); *see also* Whitehead v. Food Max of Miss., Inc., 332 F.3d 796, 807 (5th Cir. 2003) (en banc) (improper purpose to cause opposing party "embarrassment"). The author was counsel to the attorney sanctioned in the Whitehead case.

The problem is not just vagueness. It is also the apparent willingness of the courts to condemn a purpose to achieve goals that are the ordinary incidents of litigation. A person is normally presumed to intend the natural and logical consequences of his acts. Because litigation frequently causes an opponent distress, or at least inconvenience, and almost all cases settle, anyone who participates in litigation could be said to act, at least in part, with a motive to cause distress or to "pressure" an opponent into settlement. To find such purposes sanctionable invites demolition of the litigation privilege. That danger looms especially large in circuits that will sanction a pleading motivated by an improper purpose even if the pleading has legal merit and is also motivated by proper purposes.[93]

---

[88] FDIC v. Maxxam, Inc., 523 F.3d 566 (5th Cir. 2008); *see also* Morley v. Ciba-Geigy Corp., 66 F.3d 21, 24 (2d Cir. 1995) (pleading non-frivolous in some respects was sanctionable because it was "clearly an attempt to intimidate the defendant into a large settlement"); Carlton v. Jolly, 125 F.R.D. 423, 430-31 (E.D. Va. 1989) (improper for party to threaten to file a non-frivolous RICO claim if defendant would not settle). *But see* Sussman v. Bank of Isr., 56 F.3d 450 (2d Cir. 1995) (intent to force settlement not sanctionable because pleading was not frivolous); Gieringer v. Silverman, 731 F.2d 1272, 1281-82 (7th Cir. 1984) (no sanction where plaintiff who wanted a settlement filed colorably meritorious motion to intervene in class action); GEORGENE M. VARIO, RULE 11 SANCTIONS: CASE LAW, PERSPECTIVES AND PREVENTIVE MEASURES 262 & n.180, 263 (Richard G. Johnson ed., 3d ed. 2004) (collecting cases).

[89] *Maxxam*, 523 F.3d at 566.

[90] *Id.* at 587 (affirming the district court's finding of improper purpose because not clearly erroneous).

[91] *Id.* at 589.

[92] *Id.* at 583.

Examination of the courts' applications of Rule 11 carries its own uncertainties. Most of the empirical work on Rule 11 occurred before 1993 amendments that, to a degree, changed its character.[94] But some conclusions can nevertheless be drawn about its effect on peacemaking.

Peacemaker Analysis

Compel and enforce participation. The authors of Rule 11 believed that sanctions would make litigation more civilized, less costly, and more affordable. They wanted to limit disputes in court to disputes for which there was a genuine ground for disagreement about the result by punishing lawyers who file "frivolous" suits.

But insofar as participation is concerned, Rule 11's vague requirements keep out of court some potentially meritorious claims that lawyers would otherwise assert. In one survey, approximately 20% of all lawyers in the Eastern District of Washington said they had failed to assert a claim they believed meritorious because they feared Rule 11 sanctions.[95]

Rule 11 may also keep people in court who may not want to stay there. Because courts believe that Rule 11 offenses are offenses against the court, not just against the other party, the court may not accept a settlement of a Rule 11 dispute proposed by the parties.[96] They will have to fight it to the bitter end.

---

[93] *See* Marissa L. Bracke, *Where Improper Purposes Lead, Inadequate Protections Follow: Integrating the Rule 11 Improper Purpose Inquiry with the Rule 23 Protections for Absent Class Members*, 41 VAL. U. L. REV. 353, 363-66 (2006) (describing different circuit treatment of improper purpose cases); Jerold S. Solovy et al., *Sanctions Under Rule 11: A Cross-Circuit Comparison*, 37 LOY. L.A. L. REV. 727, 732-45 (2004) (describing different circuit treatment of improper purpose cases).

[94] *See, e.g.*, *supra* note 84, *infra* note 95.

[95] Gerald F. Hess, *Rule 11 Practice in Federal and State Court: An Empirical, Comparative Study*, 75 MARQ. L. REV. 313, 336, 351 & fig.32 (1992).

Respect for the parties. Either side can move for Rule 11 sanctions. Parties who do so successfully no doubt feel they have been given additional respect. But, of course, the lawyer who loses feels just the opposite, and possibly worse, because the party who moves for sanctions has been allowed to step around the litigation privilege to damage the lawyer's reputation. In one sad case, a law professor with 20 years experience filed a controversial civil rights suit, was sanctioned under Rule 11, and then became so depressed, distracted, and forgetful he walked out of a store without paying for a book, was prosecuted for stealing, and lost his job.[97]

Moreover, a Rule 11 sanctions motion threatens some lawyers—those who rely on contingent fees—more than others. In fact, they are the target of more than 70% of all sanctions motions.[98] While seemingly neutral on its face, the effect of Rule 11 in practice depends on the method by which the parties pay their lawyers. Persons who sue for damages frequently pay their counsel a fee contingent on the outcome of the case. Persons and companies who are sued most frequently pay their lawyers by the hour. A Rule 11 motion can cripple the efforts of counsel hired on a contingent fee. It forces the lawyer to invest additional time and money in the case with no hope of enhancing the ultimate contingent fee. On the other hand, a Rule 11 motion filed against counsel hired on an hourly basis can mean more hours and a higher fee for the lawyer who defends it. The disparate manner in which the two are compensated makes Rule 11 itself a powerful tool for harassment of contingent fee counsel.[99] According to survey data, nearly two-thirds of judges and lawyers think that attorneys have used Rule 11 motions to "harass and intimidate the opposition."[100]

---

[96] *See, e.g.*, ATSI Commc'ns, Inc. v. Shaar Fund, Ltd., 547 F.3d 109 (2d Cir. 2008) (refusing to vacate sanctions award even though parties requested vacatur as part of a settlement).

[97] George Cochran, *The Reality of "A Last Victim" and Abuse of the Sanctioning Power*, 37 LOY. L.A. L. REV. 691, 702-09 (2004).

[98] VARIO, *supra* note 88, at 49 ("Rule 11 as a Defendant's Tool"); Hess, *supra* note 95, at 338 & fig.24, 352 & fig.33.

[99] Stephen B. Burbank, *The Transformation of American Civil Procedure: The Example of Rule 11*, 137 U. PA. L. REV. 1925, 1937-38 (1989) (study in Third Circuit

Finally, Rule 11 can drive a wedge between the lawyer and client. If the motion is filed solely against the lawyer, and concerns the lawyer's conduct, it thrusts on the lawyer an individual interest in the litigation that is different from the client's interest. In this situation, the resulting conflict of interest may require the client to hire separate counsel and may force the lawyer to reveal communications with the client that would otherwise have been kept confidential.[101] All of this impairs the multiple ways in which a lawyer can carry out the duties that show the system's respect for the client.

Respect for the courts. Most people would agree that a court system that appears to discourage "frivolous" lawsuits and confines litigation to "proper" purposes merits respect.

But the creation of satellite litigation invites public criticism of the courts. More than one lawyer joke has been devoted to the never-ending process of litigation,[102] not to mention the more serious literature.[103] When courts sanction the intent to cause effects that are the common by-product of litigation, they invite the multiplication of satellite litigation.

Finality. Rule 11 delays finality. When the merits have been decided and the case is at last over, Rule 11 continues the fight. But instead of being a contest between the parties, it can become a contest between one party and the other party's attorney. That attorney will defend his reputation at all costs—in other words, as long and as hard as the attorney can. The parties cannot settle a Rule 11 motion without court approval, and

---

showed Rule 11 had a disproportionately adverse impact on persons who sued and their counsel, perhaps because they are the ones who use contingent fees).

[100] Hess, *supra* note 95, at 344.

[101] ANNOTATED MODEL RULES OF PROF'L. CONDUCT 107 (6th ed. 2008) (interpreting Model Rule 1.6(b)(5)); Karen S. Beck, Note, *Rule 11 and its Effect on Attorney/Client Relations*, 65 S. CAL. L. REV. 875, 901, 912 (1992).

[102] In one common joke, while two farmers who both claim ownership of a cow pull on opposite ends of the cow, their lawyers milk the cow.

[103] *See, e.g.*, CHARLES DICKENS, BLEAK HOUSE (Norman Page ed., Penguin Books 1971) (1853).

the sanctioned attorney cannot appeal a sanctions order until a final judgment has been reached in the underlying case.[104]

In particular, it goes against the Peacemaker Test and postpones finality to use Rule 11 to sanction an attorney for the "improper purpose" of taking a meritorious action to put pressure on an opponent to settle. Sanctioning an attorney for taking a legally meritorious step to motivate the other side to settle encourages delay. Ironically one "improper purpose" mentioned in the rule is delay. In fact, sanctions have been sought to punish attorneys for refusing to settle quickly enough.[105]

Reconciliation. Rule 11 battles against reconciliation. Lawyers generally deal with each other in an easy and professional manner because they know the fight is between their clients, and not between the lawyers. Lawyers on opposite sides commonly sympathize with each other about the problems their clients or witnesses present. They deprecate immaterial weaknesses in their cases to each other in polite and guarded ways. They extend courtesies to each other. They do all the things that the champions of civility urge that they do: agree to extensions of time, cooperate in setting schedules for depositions, and deal with each other in a civil manner.[106]

But a sanctions motion destroys the professional relationship.[107] Scorched earth warfare begins in earnest when one lawyer accuses another of being a legal ignoramus, or a malicious incompetent, and asks a judge to confirm that this is true in a published written opinion. In one early survey, only 8% of the judges polled thought that Rule 11 moderated contentious behavior, while almost half of the judges polled thought Rule 11 made that behavior worse.[108]

---

[104] Cunningham v. Hamilton County, Ohio, 527 U.S. 198 (1999).

[105] Nat'l Ass'n of Gov't Employees v. Nat'l Fed'n of Fed. Employees, 844 F.2d 216 (5th Cir. 1988) (failure to compromise case was not ground for sanctions when both parties made colorable arguments).

[106] Dondi Props. Corp.v. Commerce Sav. and Loan Ass'n, 121 F.R.D. 284 (N.D. Tex. 1988) (adopting standards of litigation conduct).

[107] *See* Hess, *supra* note 95, at 334 & fig.20.

[108] VARIO, *supra* note 88, at 45 (citing ELIZABETH C. WIGGINS & THOMAS WILLGING, RULE 11: FINAL REPORT TO THE ADVISORY COMMITTEE ON CIVIL RULES OF THE JUDICIAL CONFERENCE OF THE UNITED STATES § 2A, at 10-11 (1991)).

Rule 11 also makes reconciliation more difficult when a marginal Rule 11 motion is used to gain a tactical advantage in litigation. Attorneys may seek to make a motion for relief more persuasive by adding a sanctions motion to it as a rhetorical tactic. As Aristotle taught, the persuasive force of speech depends not only on the content of the speech but also on the reputation and bearing of the speaker and the emotions of the hearer.[109] The Rule 11 motion allows the attorney to attack not only the opposing party's logic, but also the opposing party's reputation and credibility.

At the same time the sanctions motion may frame the dispute so that the judge will think that something has been given to both sides if the sanctions motion is denied while the moving party's relief on the merits is granted. For these reasons, when a court looks at a Rule 11 motion, perhaps the first thing it should do is carefully examine the motives and conduct of the moving party, not those of the party against whom the motion is filed.

In summary, from a Peacemaker Test point of view, Rule 11, especially when applied to legally meritorious pleadings filed for an alleged "improper purpose," presents several problems. It can prevent some claims from being asserted, can promote disrespect for lawyers who work on a contingent fee basis, can create satellite litigation that cannot be settled, and can provoke enmity between the lawyers that will make reconciliation more difficult. It stokes the fire under the attorneys, and it should be no surprise that they sometimes unprofessionally let off steam as a result. When invoked to punish those who are trying to put a case in a position to settle, it punishes those whom the law ought to praise.

Of course, all of this might be justified if Rule 11 were necessary to protect the smooth working of the courts. But there are other remedies for lawyer misconduct. Lawyers understand that they need to protect their reputations by only filing papers that a judge will find credible. As Justice John Paul Stevens has contended, "most lawyers are wise enough to know that their most precious asset is their professional reputation. Filing unmeritorious pleadings inevitably tarnishes that asset."[110]

---

[109] *See* ARISTOTLE, *Rhetoric, in* 2 THE COMPLETE WORKS OF ARISTOTLE, 2152, 2198-99 (Jonathan Barnes ed., 1984).

In addition, a truly wronged party can sue for abuse of process or malicious prosecution if the party has been the target of a lawsuit filed for a collateral purpose or one not based on probable cause. For that, however, the wrong inflicted has to be sufficiently serious to support a separate lawsuit and a jury trial.[111] Short of that, a court can hold disobedient counsel in contempt or refer misconduct to the bar association. Bar proceedings are run by a bar prosecutor whose job is to protect the profession and the public from the slings and arrows of outrageous lawyering. Unlike the adversary in litigation, the bar prosecutor presumably will seek justice, not vengeance.[112]

Lawyer advertising

Before 1977, bar ethics rules prohibited most attorney advertising. They did so for several reasons.

First, the bar wanted to encourage attorneys to attract clients by developing a reputation for skill through the representation of others. It believed that advertising would diminish the influence of reputation. Dependence on reputation encouraged lawyers to represent clients effectively and to treat other lawyers professionally so that those clients and lawyers would refer new clients to them in the future.[113] A lawyer's dependence on reputation not only encouraged lawyers to be effective, but also helped the clients by putting their disputes in the hands of lawyers whose reputations would help them persuade a court that the client's case was just.[114] As noted above, the persuasiveness of speech depends in part on the reputation of the speaker.

---

[110] Cooter & Gell v. Hartmarx Corp., 496 U.S. 384, 413 (1990) (Stevens, J., dissenting).

[111] *See* Bus. Guides, Inc. v. Chromatic Commc'ns Enters., Inc., 498 U.S. 533, 551-52 (1991) (The Supreme Court compared Rule 11 to the common law torts and found that the more limited nature of the Rule 11 remedy meant that the adoption of Rule 11 did not violate the Rules Enabling Act, 28 U.S.C. § 2072, which says rules are not to "abridge, enlarge, or modify any substantive right.").

[112] *But see* Joy, *supra* note 84, at 791 (few referrals to disciplinary authorities are made).

[113] Note, *Sherman Act Scrutiny of Bar Restraints on Advertising and Solicitation by Attorneys*, 62 VA. L. REV. 1135 (1976) (written by the author).

Second, the bar was aware of the ease with which lawyer advertising could mislead potential clients. The economic nature of the legal services market makes it difficult for clients to assess the value of the services they get. For example, the legal market is not like the market for a car. A consumer who wants to buy a car will take a test drive, have the car inspected, return the car if it is defective, and hold out the carrot that, if the car performs well, the consumer will return to the dealership for another car when the first one wears out. The legal market knows none of this. An individual client cannot try out a lawyer's services, or demand a refund if a case is lost, or judge whether a lawyer performs well, or hold out the carrot of substantial future business.

Finally, the bar wanted to discourage litigation. Cognizant of litigation injury, and the inevitable pressures toward escalation inherent in litigation,[115] the bar wanted to save litigation as a last resort for those with pre-existing grievances and wanted to avoid "stirring up" litigation by opportunists who would not otherwise be motivated to sue.

In 1977, however, the U.S. Supreme Court held that this logic could not justify restrictions on a lawyer's First Amendment rights to engage in truthful commercial speech. In *Bates v. State Bar of Arizona*,[116] the Court interpreted "truth" broadly. It held, for example, that an advertisement stating that an attorney's fees for a name change and other services were "reasonable" was not a claim about the quality of the attorney's services even though the reasonableness of any price necessarily depends not only on the price but also on the services provided and the need for them.[117] It rebuffed a claim that the restrictions encouraged professionalism. Professionalism, it said, was a matter of etiquette, not ethics, and an anachronism whose "historical foundation[s] . . . [have] crumbled."[118] With respect to stirring up lawsuits, it commented that surely some of the lawsuits stirred up would be just and that it did not view a lawsuit as an "evil in itself."[119] In another case the Supreme Court held that an advertisement "in bad taste" was constitutionally protected.[120] Later the court flatly declared that "[a]ccess to [the] civil courts is not an evil to be regretted; rather, it is an attribute of our system of justice in which we ought to take pride."[121] It then discounted bar concerns about "dignity" in advertising by saying advertising did not take place in the courtroom, and the state could not suppress information just because some who heard it would be offended.[122]

---

[114] *See infra* pp. 35-36 (for a discussion of the importance of a speaker's reputation).

[115] Munford, *supra* note 3, at 388-89.

[116] 433 U.S. 350 (1977).

[117] *Id.* at 366. The *Bates* court rejected the thrust of the author's law review note, *see supra* note 113, by saying reputation was likely to be available only in small communities. 433 U.S. at 374 n.30. But there are just as many lawyers per capita in cities as there are in small communities, if not more. In cities, each court will attract more

But then the court, without distinguishing these cases, appeared to signal that respect for the courts might have a role to play in advertising cases. It dismissed an Iowa Supreme Court ruling, which upheld prohibitions against advertising

---

lawyers who will in turn be able to provide more information about the lawyers who practice in that court. Also, today's publishers have begun to collect and sell information about lawyer reputations using titles such as "Best Lawyers in America," or "Super Lawyers," or "Chambers USA Leading Lawyers for Business." And internet blogs and social media may also make information about a lawyer's reputation easier to obtain.

[118] 433 U.S. at 371-72.

[119] *Id.* at 376 & n.32. The advertisement before the Supreme Court in *Bates* advertised, among other things, name changes for $95 plus a $20 court filing fee. Speaking years later, plaintiff Van O'Steen said his firm created the advertisement solely for the purpose of challenging the advertising ban and ran it only one time. The firm had done only a few name changes and, after it won, the firm changed the content of its ads. Interview with Van O'Steen, Attorney for Plaintiff, Bates v. State Bar of Ariz. (July 22, 2009). *See also* Van O'Steen, *Bates v. State Bar of Arizona: The Personal Account of a Party and the Consumer Benefits of Lawyer Advertising*, 37 ARIZ. ST. L.J. 245, 247 (2005).

[120] *In re* R.M.J., 455 U.S. 191, 205-06 (1982).

[121] Zauderer v. Office of Disciplinary Counsel, 471 U.S. 626, 643 (1985). The court did acknowledge that lawsuits consume social resources and may produce "much discord and unpleasantness." *Id.* at 642. It reasoned that, if vexatious litigation were filed, the courts could impose sanctions. *Id.* at 645 n.12. That reasoning, however, overlooked the tendency that sanctions law has to cause the very "vexatiousness" it seeks to prevent. *See supra* pp. 22-30.

[122] *Zauderer*, 471 U.S. at 647-48.

claims that were not "verifiable" or were "self-laudatory,"[123] for want of a substantial federal question—equivalent to an affirmance.[124] A concurring opinion in the Iowa Supreme Court had relied on the need to have public respect for the courts in order to ensure their effectiveness.[125] In 1995, the Supreme Court approved a "black out period" that prevented Florida lawyers from sending solicitation letters to accident victims within 30 days after an accident.[126] In so doing it accepted that a state could act to prevent the "erosion of confidence in the profession."[127] The opinion, however, did not mention the Iowa case.

These intimations of constitutionality were enough to make the New York state judiciary believe in 2007 that it could restrict lawyer advertising in ways the Supreme Court has never expressly approved. Taking as its cue the need for lawyer advertising to convey truly useful facts, it said a lawyer advertisement could not "rely on techniques to obtain attention that demonstrate a clear and intentional lack of relevance to the selection of counsel."[128] It also prohibited the use of a moniker or motto that "implies an ability to obtain results in a matter."[129]

A law firm in upstate New York that depended on advertising to attract clients quickly picked up the gauntlet that the New York bar had thrown down. It admitted that its advertising violated those rules. Its advertising used jingles and special effects, like wisps of smoke and blue electrical current, to get attention. It depicted lawyers giving advice to space aliens, and employed the slogan "heavy hitters."[130] Without question it thus relied on irrelevancies to get attention and implied that it was better able than other firms to get results. The law firm sued for a declaration that the new rules violated its First Amendment rights. The United States Court of Appeals for the Second Circuit ruled in the firm's favor in *Alexander v. Cahill*.[131]

---

[123] Comm. on Prof'l Ethics v. Humphrey, 355 N.W.2d 565, 568, 571 (Iowa 1984), *vacated*, 472 U.S. 1004 (1985), *remanded to* 377 N.W. 2d 643 (Iowa 1985), *appeal dismissed & reh'g denied for want of a substantial federal question*, 475 U.S. 1114 (1986).

[124] Hicks v. Miranda, 422 U.S. 332, 344 (1975).

[125] *Comm. on Prof'l Ethics*, 377 N.W.2d at 652 (Reynoldson, C.J., concurring).

[126] Fla. Bar v. Went for It, Inc., 515 U.S. 618 (1995); *see also* Fla. Bar v. Pape, 918 So. 2d 240 (Fla. 2005) (upholding ban on use of "pit bull" image); Ralph H. Brock, *"This Court Took a Wrong Turn With Bates:" Why the Supreme Court Should Revisit Lawyer Advertising*, 7 FIRST AMENDMENT L. REV. 145, 151-69 (2009) (chronological discussion of cases governing attorney advertising).

[127] *Went for It, Inc.*, 515 U.S. at 635.

[128] N.Y. COMP. CODES R. & REGS. tit. 22, § 1200.0, rule 7.1(c)(5) (2010).

[129] *Id.* at § 1200.0, rule 7.1(c)(7) (2010).

While the court agreed that "preventing the erosion of confidence in the legal profession" was a legitimate state interest, it gave that interest no weight when it evaluated these rules. Instead, it treated the interest as being indistinguishable from the state's interest in preventing false advertising.[132] The advertisements, it said, did not serve that interest. It reasoned that the use of special effects to get attention helped the firm get its message across and did not convey anything false because no one would think that the firm actually counseled space aliens.[133] It said the motto "heavy hitters" was merely "the kind of puffery that is commonly seen, and indeed expected, in commercial advertisement generally."[134] As such, there was no proof that it had misled anyone. In so ruling the court relied on an earlier ruling by the Second Circuit that the Constitution protected a New York brewery's right to use on its label a depiction of a frog "giving the finger."[135]

Peacemaker Analysis

Participation. Traditional ways in which lawyers have made themselves known in the community generally have not increased participation in the legal system. A lawyer "advertises" when the lawyer represents clients successfully, volunteers for highly-publicized cases, runs for political office, joins bar groups and civic clubs, writes articles, sponsors a Little League baseball team, or lists the lawyer's name in the Yellow Pages. These forms of advertising assume that potential clients know when they have a grievance. They invite a client with that knowledge to seek out the lawyer who has done these things.

---

[130] Alexander v. Cahill, 598 F.3d 79, 83-84 (2d Cir. 2010), *petition for cert. filed sub nom.* Cahill v. Alexander, No. 10-203, 2010 WL 3236712 (Aug. 9, 2010).

[131] *Id.* at 83.

[132] *Id.* at 87-103 (outlining and applying the *Central Hudson* test for Constitutional limitations on commercial speech). The law firm's brief denied that dignity alone could ever be a valid state interest. Appellee's Principal and Response Brief at 44, Alexander v. Cahill, 598 F.3d 79 (2d Cir. Feb. 1, 2008) (Nos. 07-3677 & 07-3900).

[133] *Alexander*, 598 F.3d at 94.

[134] *Id.* at 95.

[135] *Id.* at 95 (citing Bad Frog Brewery, Inc. v. N.Y. Liquor Auth., 134 F.3d 87, 96 (2d Cir. 1998)); *see also Bad Frog Brewery, Inc.*, 134 F.3d at 91.

Lawyer advertising, on the other hand, generates demand for legal services. It brings more "customers" into the legal system. In fact, *Bates*' basic assumption is that advertising by lawyers will lower prices and so make it cheaper for more people to exercise their legal rights.[136] It assumes more lawsuits will make a better society because "vexatious" suits can be weeded out.[137] It overlooks that almost all litigation is "vexatious" in the sense that it inflicts litigation injury on the participants.

Unlike the advertisements at issue in *Bates*, however, the techniques questioned in *Alexander* did not do much, if anything, to bring meritorious claims into court. They said nothing about legal rights and so did not educate the public. In fact, it is doubtful that they encouraged anyone to seek legal help who would not otherwise have done so. Their primary effect may have been to attract litigants to the advertising law firm who would have otherwise chosen a different attorney.[138] If so, they did nothing to encourage participation in the legal system as a whole. The Second Circuit thought the techniques got attention,[139] but that has no peacemaking value if it simply reshuffles the deck of those already participating. And, to the extent that the techniques brought people into the system, they were only persons on the margin with grievances so minor that they would pay attention to a lawyer advertisement only when it used blue electrical current, smoke, and slogans.

---

[136] Bates v. State Bar of Ariz., 433 U.S. 350, 377 (1977).

[137] Zauderer v. Office of Disciplinary Counsel, 471 U.S. 626, 645 n.12 (1985).

[138] *But cf.* Dunagin v. City of Oxford, 718 F.2d 738, 748-50 (5th Cir. 1983) (en banc) (noting but rejecting district court finding that liquor advertising only shifted consumers between brands).

[139] Alexander v. Cahill, 598 F.3d 79, 94 (2d Cir. 2010).

Honor the litigants. The electric lighting techniques at issue in *Alexander* invite the public to view litigation as a game that has a kinship to pinball. At the end of the day, it is difficult to believe that a client attracted by those techniques would feel that the process was fairer or more just because of them. And the person who is sued would always wonder whether the techniques, rather than his own conduct, were responsible for the suit. To the extent the techniques encourage litigants to see litigation as a game, where "heavy hitting" counts and advertising tricks abound, they encourage disrespect for the system as a whole. That in turn means disrespect for those who are parties to it.

Some lawyer advertising shows respect for potential clients by leading them to better lawyers. Advertising may allow a lawyer to improve both the lawyer's skills and the lawyer's reputation by making it possible for the lawyer to concentrate in a specialized area of the law, e.g., defense of drunk driving cases or suits against railroads. Unskilled lawyers who advertise may send the clients they take on to more capable lawyers if the case goes to trial.

But lawyer advertising can also give an advantage to less-reputable and so less-effective lawyers.[140] Lawyers who depend on their professional reputations to get clients improve their reputations if they do a good job for their clients. Lawyers who advertise, however, do not need to make that effort. An ABA study found that people mistakenly think a lawyer is more highly skilled just because he has advertised, and cartoonish advertisements are especially effective in making people believe that a lawyer is more knowledgeable and intelligent.[141]

In fact, even law office efficiency may prove to be problematic from the client's point of view. Praise for "low cost" legal services overlooks that many of the clients advertising attracts do not pay by the hour but instead pay by giving a contingent fee to the lawyer. In contingent fee cases, a "low cost" lawyer is not necessarily the one who will obtain the best result for the client. The lawyer with lower costs can achieve a desired hourly return on his effort with a lower settlement than a lawyer with higher costs would require.[142] A study of "settlement mills" run by lawyers who advertised heavily found some benefit for clients with small claims but discovered harm to clients with large claims because those clients got less money when they took their case to a settlement mill firm.[143]

---

[140] *See supra* note 113.

[141] William E. Hornsby, Jr. & Kurt Schimmel, *Regulating Lawyer Advertising: Public Images and the Irresistible Aristotelian Impulse*, 9 GEO. J. LEGAL ETHICS 325, 353 & tbl.2 (1996).

The *Alexander* facts present a further problem for the potential client. The Second Circuit said the "heavy hitters" slogan was not misleading because it was "puffing," i.e., a statement of opinion that no one would believe.[144] But the lawyers who used it would not have done so if it had not attracted clients by implying an ability to get results—that was why it ran afoul of the New York rules. More generally, while some statements of opinion cannot be proven true or false, an opinion can imply the existence of false facts.[145] That is especially true where, as in *Alexander*, no supporting facts are given. And it is false to offer the conclusory opinion that any lawyer is a "heavy hitter;" whether or not a lawyer is a "heavy hitter" in a case depends on many factors, including the jurisdiction, the judge, the jury, and the skill of opposing counsel. A lawyer who is a "heavy hitter" in one setting will be a "weak sister" in another. To this extent, those who come to the law firm because of the use of the slogan "heavy hitter" are not well served.

---

[142] For example, assume a 40% contingent fee. A lawyer with $200 in overhead who wants to net $200 for an hour spent on the case can do so if the hour produces a $1,000 settlement. That would leave $600 for the client and $400 for the attorney, from which he pays his $200 expense. But another lawyer with $400 in overhead, who wants to net $200 for an hour spent on the case, would need to demand a $1,500 settlement. That would leave $900 for the client and $600 for the lawyer. In contingent fee cases, the client can be better off when his attorney has a higher overhead. Generally speaking, negotiators with higher goals get better results. G. RICHARD SHELL, BARGAINING FOR ADVANTAGE: NEGOTIATION STRATEGIES FOR REASONABLE PEOPLE 28-32 (2d ed. 2006).

[143] Nora Freeman Engstrom, *Run-of-the-Mill Justice*, 22 GEO. J. LEGAL ETHICS 1485 (2009).

[144] *See supra* note 134 and accompanying text; *see also* Pizza Hut, Inc. v. Papa John's Int'l, Inc., 227 F.3d 489, 496 (5th Cir. 2000).

[145] *See* Milkovich v. Lorain Journal Co., 497 U.S. 1, 18 (1990).

Command respect from the litigants and the public. Bar association attempts to ensure that lawyer advertising is "dignified" promote public respect for the justice system and help preserve the ability of the courts to make peace in society. What litigants want most is a "dignified, careful and unbiased hearing at their case."[146] If justice were made by machines, then perhaps it could be advertised like a vegematic, the does-anything gadget that was once prominent in late-night television sales pitches. But justice is human. To make peace effectively, courts need the public respect, and even the public awe, that the word "dignified" captures.

Peacemaking is not just a good idea. Rather it is a constitutional value just as representative government and free speech are constitutional values. A civilized and free society cannot exist without a basic peace among its citizens. The Constitution's preamble sets out its purposes, and one of those is to "ensure domestic tranquility."[147] The courts can ensure tranquility only if they are respected. The power of courts depends on public respect for the law.[148] It is hard to believe that the *Alexander* opinion would be equally respected if it came from randomly selected derelicts in t-shirts and Bermuda shorts who sat at a card table in New York's Foley Square and scribbled in long-hand.

Lawyers who advertise using electric flashes, space aliens, and blue smoke do not promote respect for the courts. "The state courts cannot be disassociated, in the minds of lay persons, from the conduct of their officers, either in or out of court."[149] When attorneys act like clowns, the lay person is bound to think that the courtroom is a circus.

The peacemaking role of the courts, and lawyers as officers of the courts, in turn distinguishes lawyer advertising from all other commercial advertising. A brewery does not need respect to sell beer, but a court must have it in order to make peace effectively. This distinction between courts and businesses is reflected in the distinction found in the law of contempt. There is no law against expressions of contempt for commercial institutions, e.g., beer manufacturers. But the law forbids statements that show contempt of court.[150]

---

[146] *See* E. ALLEN LIND ET AL., *supra* note 49, at 45.

[147] *See* Munford, *supra* note 3, at 380-81.

[148] *See* ALEXIS DE TOCQUEVILLE, I DEMOCRACY IN AMERICA 151 (Phillips Bradley ed. 1945) (discussing the importance of the U.S. Supreme Court).

[149] Comm. on Prof'l Ethics v. Humphrey, 377 N.W.2d 643, 649 (Iowa 1985) (Reynoldson, C.J., concurring).

Finality. Lawyer advertising would not seem to have much to do with the finality of a dispute except to the extent that it brings into court disputes that otherwise might have been resolved out of court. This is no different from the issue of participation previously discussed.[151]

Reconciliation. If lawyer advertising settles pre-existing disputes more quickly, it promotes reconciliation. On the other hand, to the extent it creates disputes, creates unjustified expectations, or unfairly damages the reputations of those the lawyer would like to sue, it prevents reconciliation.

From a peacemaker point of view, it is doubtful that the lawyer advertising techniques at issue in *Alexander* brought anyone with a meritorious claim into court who would not have been there already. The techniques did nothing to honor those already there. Even worse, the legal advertising techniques approved in *Alexander* are misleading and threaten the respect the court system needs to function as an effective peacemaker. The use of phrases like "heavy hitters" not only implies false facts but also suggests that a lawyer will get results through physical intimidation. Like the motto "pit bull," which was banned in Florida, it makes a "mockery of our dedication to promoting public trust and confidence in our system of justice."[152]

Bar associations should be allowed to regulate advertisements that use attention-getting techniques irrelevant to the practice of law or use slogans that imply an ability to get results. Those techniques undermine the respect the courts must command if they are to carry out their constitutional peacemaking role. Any message worthy of being communicated can be expressed without them. Also, there are many more respectful ways of attracting attention through advertising that do not involve wisps of smoke, blue electrical current, talking to space aliens, or a misleading motto.

---

[150] *See* Carroll v. Jaques, 926 F. Supp. 1282 (E.D. Tex. 1996) ($1,000 fine for telling opposing lawyer in deposition "f... you"); State v. Blackwell, 2007 WL 4258336, at *3 (N.J. Super. Ct. App. Div. 2007) (defendant held in contempt for "flipping the finger" at a prosecutor).

[151] *See supra* pp. 35-36.

[152] Fla. Bar v. Pape, 918 So. 2d 240, 246 (Fla. 2005).

CONCLUSION

Lon Fuller listed aspirational prerequisites that would enable persons affected by a law to agree on what it means and act accordingly. His list included, for example, requirements that the law be found in public rules that were understandable and did not change.[153]

The five questions the Peacemaker Test asks expand on Fuller's list. As demonstrated by the examples above, peacemaking also requires control over who participates in the dispute, respect for the parties, public and party respect for the peacemaker, finality, and steps toward reconciliation. These are the "social processes" that constitute the "reality" of dispute resolution that Fuller hoped would someday be identified.[154]

The Peacemaker Test does not provide a trump card. It only supplies one way to evaluate legal rules. It competes with economic efficiency, deference to the legislature, religious values, and critical legal studies, among other types of analysis. But it is a card that should be played if we are to enjoy domestic tranquility in its fullest measure.

---

[153] FULLER, *supra* note 1, at 39; *see also* Martin P. Golding, *Jurisprudence and Legal Philosophy in 20th Century America: Major Themes and Developments*, 36 J. LEGAL EDUC. 441, 474 (1986) (describing Fuller's philosophy as "natural law in its 'most modest form'").

[154] *See supra* text accompanying note 1.

ESSAY

A STATE OF RIGHTLESSNESS: THE EGREGIOUS CASE OF NORTH KOREA

Morse H. Tan[*]

I. INTRODUCTION

In the midst of the important discussion of geopolitical security concerns regarding nuclear weapons, the stifled cry of the North Korean people, whose rights have been grossly and systematically trampled upon, pleads for redress. The egregious conditions and unthinkable treatment of the North Korean people by their own government exclaims for worldwide attention. The dictatorships of Kim Jong-Il and his father Kim Il-Sung who reigned before him have defied treaties entered into throughout the international community and have made decisions that focus purely on the gain of the government to the extreme detriment of its citizens.

Media attention regarding these violations has been weak albeit improving. For instance, the United States news sources appropriately informed us about the plight of American journalists in North Korea (who have since been returned); yet not much attention focuses upon the astonishing absence of a free press, free speech, and free-association rights inside North Korea. Media outlets readily cover North Korea's missile launches, but do not provide such coverage for the concentration camps that can stun even the calloused of heart. The story of North Korean counterfeiting U.S. currency makes for a fine news headline, yet the wrenching narratives of refugees who are sent back to North Korea to face cruel torture, unrelenting labor on less than subsistence rations, and possible execution receive all too little recognition.

---

[*] Associate Professor, Florida Coastal School of Law. Prof. Tan wishes to thank his student assistants: Branden Vicari, Frank Newkirk, Lisa DiFranza, Alicia Magazu, Caroline Kubovy, and Simon Trautmann for their work in helping to bring this essay into existence. Branden, Lisa and Alicia in particular deserve special recognition for their work. Thanks to Scott Luftglass for his encouraging correspondence. This essay is dedicated to my late, paternal great grandfather, who served as Governor of Pyongyang Province of a then unified Korea. As I imagine that great grandfather would desire, may the future bring a peacefully united Korea that upholds and respects human rights.

The story of Robert Park briefly increased international awareness of human rights violations in North Korea. Park, a courageous man who sought to increase worldwide attention to the human rights atrocities in North Korea, knowingly put himself at risk by walking into North Korea with requests for Kim Jong-Il and other top leadership to desist from their human rights abuses and step down from power. After enduring severe abuse in prison, he was released in February 2010 after spending forty-three days in North Korean custody.[1]

This essay steps into the relative dearth of popular and legal academic treatment to analyze this egregious state of rightlessness (an intentional neologism) and concludes with reflections upon possible judicial redress options. Experts in the human rights field have averred that the human rights situation in North Korea is the worst in the world.[2] The North Korean government denies any human rights abuses, insisting "that 'there is no human rigths [sic] problem in North Korea.'"[3] This academic work challenges this outright denial by providing an analysis of various human rights abuses persisting north of the most heavily armed border in the world.

---

[1] Adam Gabbot, *U.S. Human Rights Campaigner Freed by North Korea Returns Home*, Guardian, Feb. 7, 2010, http://www.guardian.co.uk/world/2010/feb/07/robert-park-north-korea-home.

[2] *See generally* David Scheffer & Grace Kang, Editorial, *North Korea's Criminal Regime*, N.Y. TIMES, July 7, 2006, http://www.nytimes.com/2006/07/06/opinion/06iht-edkang.2131969.html; Stephanie Ho, VOA News, *Freedom House Calls for Action on North Korean Human Rights Violations*, US FED NEWS, June 6, 2007. North Korea is also included in the report by Freedom House titled *The Worst of the Worst 2010: The World's Most Repressive Societies*, *available at* http://www.freedomhouse.org/uploads/special_report/88.pdf.

[3] Sung-Chul Choi, *Human Rights in North Korea in the Light of International Covenants*, *in* INTERNATIONAL COMMUNITY AND HUMAN RIGHTS IN NORTH KOREA 53, 56 (Sung-Chul Choi ed., 1996) (quoting a statement by Nam-Joon Paek in 1993 at the World Conference on Human Rights in Vienna).

North Korea epitomizes and seems to justify the "realist" view of international relations in terms of how it operates. As such, North Korea acts as if neither international law, nor the international treaties it has already ratified, apply to it. *Pacta sunt servanda* does not constitute North Korea's *modus operandi.* Furthermore, North Korea does not allow itself to be accountable to any other country or international organization; rather, it clings only to its own oppressive sovereignty.

Along these lines: "North Koreans claim that, 'Human rights is [sic] unthinkable apart form [sic] national independence, and what imperialists call rights is their power to do whatever they please; thus, the biggest enemies of human rights are the imperialists who intervene in other states' internal affairs under the pretense of protecting human rights.'"[4] The claim that human rights is unthinkable is accurate when put in another context: the current state of rightlessness for many North Koreans causes a bleak sense that under the present government, human rights do seem unthinkable because the government has deprived them of their most basic human rights. The accusation against "imperialists" of "doing whatever they please" actually provides an apt description of Kim Jong-Il, the self-styled Dear Leader, who does whatever he pleases. Unfortunately, doing whatever he pleases for Kim Jong-Il and his "elite" includes massive and egregious violations of human rights, especially against those in the country he rules with an iron fist.

Contrary to the constant denial of the North Korean government, this essay seeks to help establish the case for this state of rightlessness, the egregious case of North Korea. It lays the foundation for a subsequent article regarding judicial redress of these gross and systematic violations of human rights. Before any need for judicial redress raises itself for consideration, the case that cries out for such a forum must receive delineation: this essay proposes to expound on the major violations of human rights in North Korea together with the context that makes such violations possible.

---

[4] *Id.* at 55 (quoting Kim Jong Il, *Socialism is a Science*, RODONG SHINMOON [THE LABOR NEWSPAPER], Nov. 4, 1994).

II. A STATE OF RIGHTLESSNESS

A. The Putative Domestic Law

A cursory examination of North Korea's constitution might insinuate that it gives various rights and freedoms. However, a closer reading reveals that these purported rights and freedoms prove illusory. All of these rights and freedoms are nullified by additional terms in this same constitution. This nullification takes place by subordinating all such rights and freedoms to the will of the Dear Leader, Kim Jong-Il, and the advancement of the ruling party.

In its ongoing propaganda campaign, perhaps in response to rising international pressure to respect the human rights of its own people, North Korea revised its constitution in April 2009 and made it available to the public in late September 2009.[5] The new constitution purports to grant the people a right to elect and recall an assemblyman and declares that the people are the "masters of everything."[6] Of course, one should not ignore the defunct legislature, the totalitarian dictatorship, and concentration camps full of slave labor in straining to find any substantial meaning in actual praxis to these provisions.

The current constitution declares that the State shall respect and protect the human rights of the working people.[7] This provision has never been included in any prior constitution of North Korea.[8] However, these new constitutional provisions merit a massive measure of skepticism.[9] Testimony has revealed public executions for "anti-regime" charges (e.g. a sergeant going AWOL), for economic crimes (e.g. theft of government property or theft of livestock), and for socially delinquent behaviors (e.g. streaking or disco dancing).[10] Unfortunately, such extremely expansive utilizations of capital punishment are not hypotheticals, but actual instances. So unless the government has engaged in a tectonic shift, a rather improbable prospect, comparable actions are likely continuing, notwithstanding the new language in the current constitution.

---

[5] Choe Sang-Hun, *New North Korean Constitution Bolsters Kim's Power*, N.Y. TIMES, Sept. 29, 2009, at A12, *available at* http://www.nytimes.com/2009/09/29/world/asia/29korea.html.

[6] Northeast Asia Matters, http://asiamatters.blogspot.com (Oct. 8, 2009).

[7] *Id.*

[8] Sang-Hun, *supra* note 5.

[9] Lee Keum-Soon et. al., *White Paper on Human Rights in North Korea*, KOREA INST. FOR NAT'L UNIFICATION 97 (2004) [hereinafter KINU 2004].

The constitution purports to expand the already absolutist authority held by the Dear Leader, who currently serves as the chairman of the National Defense Commission (NDC), head of the Workers' Party, and as the Supreme Commander of the People's Army. Chapter VI Article 100 of the constitution pronounces the NDC chairman the supreme leader of the Democratic People's Republic of Korea (DPRK)—a reaffirmation of the already existing political structure.[11] Moreover, a continued reading exposes these purported constitutional rights for the people of the DPRK as illusory since the NDC may "[r]escind the decisions and directives of state organs that run counter to the orders of the chairman of the DPRK."[12] In other words, Kim Jong-Il's orders trump any right supposedly given by the constitution. While the new constitution may have revised some provisions, a yawning chasm between the putative law and its implementation has persisted.

North Korea has revised its criminal code four times as of 1999, maintaining the heaviest penalties for those who injure the authority of the two Kims.[13] In one instance, a young child and his family were shipped to a concentration camp for the "crime" of scribbling on a picture of Kim Jong-Il with a crayon.[14]

---

[10] *Id.* at 85-89.

[11] Northeast Asia Matters, *supra* note 6.

[12] *Id.*

[13] KINU 2004, *supra* note 9, at 108.

[14] David Hawk, *The Hidden Gulag: Exposing North Korea's Prison Camps*, U.S. COMMITTEE FOR HUMAN RIGHTS IN N. KOREA (2003), *available at* http://www.hrnk.org/HiddenGulag.pdf.

In a report to Amnesty International, North Korea insisted that citizens enjoy equal rights and are "not discriminated against for reasons of race, color, sex, language, religion, political beliefs or opinions, national or social origin, property, birth or status."[15] However, there is no dearth of examples that prove otherwise. For instance, the law "strictly classifies every individual by his or her family background (or class origin) and by the degree of loyalty to the regime."[16] This practice is contrary to the assertions made in the report to Amnesty International.

The absence of an effective, politically independent judiciary also renders individuals' rights and freedoms unprotected. Furthermore, it is commonplace for those accused of political and certain economic crimes to be denied due process of the law, such as the right to an attorney for their defense.[17] This is a denial of due process because the role of attorneys in the North Korean criminal justice system is solely to protect the Worker's Party's policies instead of any particular individual's rights.[18] Trials, if they take place at all, often are superficial mockeries of the idea of due process. They often lead to summary sentencing without any semblance of a full judicial process, much less anything approximating a just result on the merits.

B. What International Law Applies

North Korea has made itself a party to important human rights treaties: the International Covenant on Civil and Political Rights; the International Covenant on Economic, Social and Cultural Rights; the Convention on the Rights of the Child; and the Convention on the Elimination of All Forms of Discrimination Against Women. It also has ratified the Biological Weapons Convention, tangentially pertinent because the use of these weapons could violate human rights.[19]

---

[15] Soo-am Kim et. al., *White Paper on Human Rights in North Korea*, KOREA INST. FOR NAT'L UNIFICATION 116 (2007), *available at* http://www.kinu.or.kr/kinu/sc/skin/kinu/data/file/data04/data/hrn2007.pdf [hereinafter KINU 2007].

[16] *Id.*

[17] KINU 2004, *supra* note 9, at 110.

[18] KINU 2007, *supra* note 15, at 111-12.

Although North Korea has ratified these treaties, it maintains its own "style" of human rights through its view that the international community's demand for improvement is an infringement on its national sovereignty. Yet the act of ratifying these treaties requires a certain degree of relinquishment of sovereignty. However, North Korea does not enter into treaties with the intention of being bound by them, but rather it uses them as leverage for tactical gain.[20] On October 21, 1994, North Korea signed the Agreed Framework, which required it to transition from graphite-moderated nuclear reactors to light water reactors over the next ten years in exchange for U.S. commitments to replace lost generating capacity through shipments of heavy oil.[21] Thereafter, U.S. intelligence discovered that North Korea had breached the Agreed Framework by continuing to develop and possess nuclear weaponry and therefore suspended U.S. shipments of heavy oil and scrapped the Light Water Reactor project.[22] North Korea then claimed the U.S. was the party who breached the agreement, and North Korea needed to develop and produce nuclear weapons to deter American aggression, which the American government has never explicitly threatened.[23] Thus, North Korea never intended to be bound by the Agreed Framework, but rather used the negotiation process as leverage to gain heavy oil.

North Korea has rejected humanitarian aid it may not control from the international community, despite increased food shortages and economic hardships the country continues to face. Consider this account as an example, "As the number of defectors increased amid food shortages and economic hardship and as international concerns over the human rights situation in North Korea mounted, North Korea has seemed to worry more about the possibilities of the international community's humanitarian intervention to fulfill the duties of human rights protection."[24] As a result, in spite of the dire shortages, North Korea has resorted to acts such as ejecting humanitarian relief organizations, including the UN World Food Programme.

---

[19] It is already known that North Korea engages in horrific biochemical experimentation in their concentration camps. *See* Robert Windrem, *Death, Terror in N. Korea Gulag*, MSNBC, Jan. 15, 2003, http://www.msnbc.msn.com/id/3071466.

[20] Morse Tan, *The North Korean Nuclear Crisis: Past Failures, Present Solutions*, 50 ST. LOUIS U. L.J. 517, 518 (2006).

[21] *Id.* at 531-32.

[22] *Id.* at 532-33.

[23] *Id.* at 533.

In 1946, the United Nations' Economic and Social Council (ECOSOC) established the Commission on Human Rights in accordance with Article 68 of the UN Charter.[25] Soon after, the committee completed a draft of the Universal Declaration of Human Rights. Although this document was unanimously adopted by the General Assembly in December of 1948, it was not initially intended to have binding force, but rather as a "common standard of achievement for all peoples and all nations."[26]

After passage, the Assembly called on "[m]ember countries to publicize the text of the Declaration and 'to cause it to be disseminated, displayed, read and expounded principally in schools and other educational institutions, without distinction based on the political status of countries or territories.'"[27] Even though the document was not binding, it was still a significant milestone in that it significantly advanced the actualization of human rights, and has been likened to widely hailed documents like the Magna Carta.[28] The document took more than two years to draft and generated a multitude of debates. Each clause of the document resulted in over 1400 votes from fifty-eight member States. In the end, it was unanimously adopted on December 10, 1948.[29] The Declaration has been translated into approximately 250 languages and is the most cited document on human rights in the world.[30] The Declaration has inspired more than sixty "legally binding treaties" designed to promote the protection of human rights.[31] Unlike any other instrument of its type, the Universal Declaration holds a special normative place in the international community.[32]

---

[24] KINU 2004, *supra* note 9, at 18.

[25] Choi, *supra* note 3, at 61.

[26] United Nations, *Universal Declaration of Human Rights*, http://www.un.org/en/documents/udhr/ (last visited Oct. 30, 2010).

[27] *Id.*

[28] Choi, *supra* note 3, at 62.

[29] *Id.*

[30] United Nations, *A United Nations Priority: Universal Declaration of Human Rights*, http://www.un.org/rights/HRToday/declar.htm (last visited Jan. 29, 2010).

The Declaration states, "[E]veryone is entitled to all rights and freedoms set forth in this Declaration . . . and to principles of equality."[33] Nonetheless,

> [T]here were states that questioned the legally binding force of this declarational document which proposes human rigths [sic] guarantees. Because of this, the 21st UN General Assembly adopted the December 16, 1966, international human rights covenants in order to provide the necessary instruments to strengthen the binding force of the previous declaration and thereby address the problem.[34]

Over the next twenty-odd years came numerous international human rights covenants along with international agreements with binding force.[35] North Korea is a party to the two primary covenants emerging in the wake of the Declaration and the U.N. Charter, the International Covenant on Economic, Social and Cultural Rights, and the Covenant on Civil and Political Rights.[36]

A number of legal scholars have argued for enhanced status for the Universal Declaration. These arguments include claims that the Universal Declaration (1) has achieved customary international law status, at least for certain rights,[37] (2) is binding as the legal exposition of UN Charter references to human rights,[38] or (3) is *the* authoritative statement of human rights in international law as a general principle of law.[39]

---

[31] Peter Van Dijk, *The Universal Declaration is Legally Non-Binding; So What? in* REFLECTIONS ON THE UNIVERSAL DECLARATION OF HUMAN RIGHTS: A FIFTIETH ANNIVERSARY ANTHOLOGY, 108, 109 (Barend van der Heijden & Bahia Tahzib-Lie eds., 1998).

[32] *Id.*

[33] Choi, *supra* note 3, at 63.

[34] *Id.* at 65.

[35] *Id.* The covenants include the following: International Covenant on Economic, Social and Cultural Rights, International Covenant on Civil and Political Rights, Optional Protocol to the International Covenant on Civil and Political Rights, Second Optional Protocol to the International Covenant on Civil and Political Rights Aimed at the Abolition of the Death Penalty, Covenant Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, Convention on the Elimination of All Forms of Discrimination against Women, Convention on the Rights of the Child, International Covenant on the Elimination of All Forms of Racial Discrimination and Convention on the Prevention and Punishment of the Crime of Genocide. *Id.*

[36] *Id.* at 65-66.

If an individual or group wishes to file a report regarding a human rights violation, they must follow the UN Human Rights Committee's Procedure 1503. When the Committee receives the report, it requires the accused state to reply and discuss the allegations in a closed session. "[T]he committee may take actions such as examining the conditions wherein the human rights violation occured [sic], appointing special inspectors and proposing mechanisms for improvements by holding open discussions or setting up temporary special committees, sending in experts, publicizing the human rights violation, etc."[40] For failure to comply with the "requirement to reply," the report of the state's human rights violation is publicized in the commission's annual report.[41] The state is supposed to be pressured by the international community's public opinion to amend the situation.

As a result of a report filed against North Korea for human rights violations, the UN assigned a special rapporteur to investigate the country and report the findings to the UN General Assembly and UN Commission on Human Rights (UNCHR). Vitit Muntarbhorn, a law professor from Southeast Asia, has fulfilled this role from remote as North Korea has not yet allowed him to extend the investigation directly within its borders. Nevertheless, Professor Muntarbhorn did produce a report that diplomatically described the adverse human rights situation in the DPRK. Specifically, he reported the North Korean government's use of guilt by association to control their people, in addition to continuing violations of "guaranteed" freedoms of thought, religion, expression, and travel.[42] In his reports to the sixty-first and sixty-second meeting of the UNCHR, Professor Muntabhorn called for the following:

---

[37] RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE U.S. § 702 (1987). This source includes the following violations of human rights as customary international law without claiming comprehensiveness: murder or causing the disappearance of individuals, genocide, slavery, torture or other cruel inhuman or degrading treatment or punishment, prolonged arbitrary detention, systematic racial discrimination and consistent patterns of gross violations of internationally recognized human rights.

[38] Thomas Buergenthal, *International Human Rights Law and Institutions: Accomplishments and Prospects*, 63 WASH. L. REV. 1, 10 (1988).

[39] Bruno Simma & Philip Alston, *The Sources of Human Rights Law: Custom, Jus Cogens, and General Principles*, 12 AUSTL. Y.B. INT'L L. 82 (1992); *see also* Jonathan I. Charney, *Universal International Law*, 87 AM. J. INT'L L. 529, 549 (1993).

[40] Choi, *supra* note 3, at 73-74.

[41] *Id.* at 74.

> (1) North Korea must fulfill the terms of all human rights treaties it has signed; (2) North Korea is encouraged to accommodate various recommendations contained in the special rapporteur's reports to the UN General Assembly; (3) South Korea should continue its support for the North Korean defectors; (4) the two Koreas should resolve the issue of abducted South Koreans; (5) North Korea should guarantee on-site observations of the distribution of humanitarian assistance materials; and (6) North Korea should include human rights improvement measures in its economic development plans.[43]

In response to the reporting of grave violations of human rights and the numerous practices of inhuman treatments such as tortures, beatings, rapes, starvation deaths, and murders of newborns, the UNHRC in its "concluding observations" strongly recommended North Korea allow onsite inspections of various correctional and detention facilities in North Korea by independent national and international teams.[44] However, North Korea said it could do it itself, and refused this proposal.[45]

C. Why There Is No Rule of Law

The North Korean regime effectively defines human rights into oblivion. "North Korea classifies its populace into the people and the enemy, and it argues that human rights mean the exercise of absolute dictatorship over the class enemies in order to build a socialist society."[46] Somehow, the crushing of perceived "enemies" of the state is rationalized as upholding human rights. An absolute, totalitarian dictatorship that regularly oppresses its populace must be considered antithetical to the rule of law and recognition of human rights, not as supportive of human rights. "Under these circumstances, no one can insist on individual freedoms and rights (such as the freedom of speech, press, religion, association and political participation) unless one is prepared to risk personal safety. It is reported that those who ask for individual rights and freedoms often disappear without a trace."[47] The deprivation of these human rights cannot be considered anything remotely resembling an embrace of human rights.

---

[42] Hawk, *supra* note 14, at 77.

[43] KINU 2007, *supra* note 15, at 12.

[44] KINU 2004. *supra* note 9, at 107.

[45] *Id.*

North Korea claims a "style" of human rights, referred to as "our style of human rights concept."[48] Unfortunately, the policies of this "style" clash completely with respect for human rights in that it has strong political tendencies: every citizen's life is subject to political necessity. The Constitution of North Korea itself states it is subordinate to the supreme dictator's ideology.[49]

> The "Socio-political organism" metaphor, which is at the core of North Korean political theory, is the element that makes individual human rights guarantees even more difficult. This is because an individual may not act outside of the guidance of the Great Leader and the Party under the philosophy that establishes a one-leader (Workers) party which combines with the masses to form a socio-political organism . . . . This also means that the Great Leader is seen as the center and "founder of life."[50]

Anything even mildly contravening, non-conforming or seen as undermining this totalitarian state has no place in the "socio-political organism."

---

[46] *Id.* at 8.

[47] *Id.* at 10.

[48] Choi, *supra* note 3, at 98.

[49] *Id.*

[50] *Id.* at 94.

The absence of an effective, politically independent judiciary also renders these rights and freedoms unprotected. As previously stated, trials, if they take place at all, often are superficial mockeries of the idea of due process, often leading to summary sentencing without any semblance of a full judicial process, much less anything approximating a just result on the merits.

D. The Cult of Kim Jong-Il

The entire nation finds itself under the thoroughgoing sway of a single individual: Kim Jong-Il. North Korea has only had two rulers, Kim Il-Sung, and his son, Kim Jong-Il. A son of Kim Jong-Il, Kim Jong-Un, reportedly looks to continue this hereditary dynasty.[51]

The state religion of North Korea is the apotheosis, the deification of the only two individuals who have ruled North Korea: the deceased Kim Il-Sung, referred to as the Great Leader and his son Kim Jong-Il, called the Dear Leader. The people of North Korea are submerged under a deluge of propaganda and every manner of cultural influence to this effect such that the entire society is oriented around exaltation of the two Kims.[52]

The press is not allowed to engage in any criticism of the leadership or the instructions of Kim Jong-Il.[53] "[T]he North Korean press disregards the proper function of the press, such as providing critical commentary and providing objective information to citizens, but instead focuses on the propaganda of Kim Il-Sung based on the Juche ideology and upon indoctrinating the population."[54] "All radio dials are fixed to the DPRK official broadcasting service channels and sealed . . . . [I]f a seal is found broken[, the person is] . . . treated as a political criminal."[55] "Nothing else may be worshipped other than Juche ideology and its founder, the Great Leader . . . . The worship of Kim Il-Sung and Kim Jong-Il is specifically stipulated in the Ten Great Principles of Unique Ideology."[56]

---

[51] Choe Sang-Hun and Martin Fackler, *Kim Jong-un*, N.Y. TIMES, Sept. 29, 2010, http:// topics.nytimes.com/topics/reference/timestopics/people/k/kim_jongun/index.html.

[52] *See generally* Choi, *supra* note 3, at 126-30

[53] *See id.* at 142.

[54] *Id.*

[55] *Id.* at 142-43.

The people, starting from childhood, sing songs praising the Kims; a woman was taken to a concentration camp for singing a South Korean pop song instead.[57] The people are taught to thank the Kims for any food that they might have, or any other positive feature, if any, in their lives. When a doctor (even a foreign doctor) performs surgeries to restore eyesight, the first thing that people do, in accordance with the massive conditioning and draconian pressures they have been subjected to along these lines, is to thank the Kims.[58] A whole web of legends is spun about the Kims that attribute superhuman abilities and incredible phenomena surrounding their births and their lives. Notwithstanding the cascade of propaganda, the people still know about the immoral life of Kim Jong-Il, which cannot be completely concealed through the State produced mélange of fabrications.

This false apotheosis does not give room for human rights and freedoms. "Under the unitary ruling system and the deification efforts for the supreme leader, North Korea is predisposed to regard any expansion of individual rights and freedoms as a challenge and potential danger to the government and the leader."[59] Indeed, North Korea "forc[es] people even to sacrifice their lives to save portraits of Kim Il-Sung and Kim Jong-Il in accordance with the Ten Great Principles."[60] So, in the event of a fire at their residence, a family is to safeguard portraits of the Kims before anything else, even if it means risking their lives. The cult of the Kims thoroughly deprecates the rights and freedoms of the bulk of their people and reserves such freedom and rights for the corpulent despot who rules over an emaciated populace and an ever-increasing pile of corpses through cruelty and neglect.

---

[56] *Id.* at 161.

[57] Hawk, *supra* note 14, at 46.

[58] *Inside: Undercover in North Korea* (Nat'l Geographic television broadcast Jan. 18, 2009), *available at* http://natgeotv.com/asia/listings/weekly/ngc/170410/3.

[59] KINU 2004, *supra* note 9, at 16.

[60] *Id.* at 163.

E. A Thoroughly Totalitarian Regime

This rulership accurately merits the use of the adjective totalitarian as merely a descriptive term. Kim Jong-Il is considered effectively above the law within North Korea, and violates it with utter domestic impunity. In effect, the will and whim of the Dear Leader substitutes for rule of law. The North Korean Constitution says it guarantees a total of 17 rights, but they in fact play only supporting roles for the 6 duties imposed on its citizens.[61] Article 75 newly added citizens' rights to free travel and residence.[62] "The most important duty for all North Koreans in their daily lives is the duty of absolute loyalty and obedience to the supreme leader, or Suryong."[63] In other words, allegiance to Kim Jong-Il trumps everything else, including human rights. "Thus, even though the basic rights listed in North Korea's constitution may be similar to those of free democratic countries, the document's fundamental purpose is the strict regimentation of the people. The purported list of the people's basic rights may be interpreted to be solely for external/internal propaganda."[64] The unimplemented rights claimed by the Constitution are a patina over a ghastly and morally moribund state.

The society of North Korea is under the control of a supreme leader—not based on institutional means of governance.[65] Consequently, to even maintain power, a totalitarian leadership such as North Korea must exert considerable effort to force allegiance and loyalty to the leader.[66] Through classification, North Korean citizens become labeled and control of the people is effectuated with efficiency.[67] Decades ago, the North

---

[61] *Id.* at 10.

[62] Northeast Asia Matters, *supra* note 6.

[63] KINU 2004, *supra* note 9, at 10.

[64] Choi, *supra* note 3, at 97.

[65] KINU 2004, *supra* note 9, at 113.

[66] *Id.* at 6.

[67] *Id.* at 139.

Korean government supplied goods and services at minimal or no cost to the people, but in recent times, such supply functions have diminished drastically.[68] North Korean citizens are denied control over their basic needs, including food, clothing, and shelter because the Great Leader controls the supplies of all. Thus, even though North Korea's constitution declares the state shall provide food, clothing, and housing to all the working people,[69] there exists a major gap between the law and reality in North Korea today.[70]

The North Korean regime controls and distributes food to its citizens through a "Food Supply System" which essentially provides food to those who work but eliminates those who do not. The idea of such a system is to encourage citizens to be obedient, thankful, and loyal to their ruler. North Korea also limits access to clothing by maintaining a low supply and inflating prices.[71] A citizen's clothing supply is dependent on his or her political status: thus, laborers receive free work clothes once or twice a year, and students and engineers receive discounted formal clothing less frequently.[72] Even though the non-elite citizens in North Korea are already unable to afford clothing, the government has further tightened its grasp on citizens by controlling housing. Every house in North Korea is "the haven of the state"; therefore, there is no recognition of any individual ownership and housing is provided to "the people from the best up to the 5th degree."[73] The "5th degree" references the categorization of the populace into over fifty categories based on political loyalty.

---

[68] *Id.* at 15-16.

[69] Northeast Asia Matters, *supra* note 6.

[70] KINU 2004, *supra* note 9, at 10.

[71] Choi, *supra* note 3, at 103.

[72] *Id.*

[73] *Id.* at 104.

III. THE EGREGIOUS CASE OF HUMAN RIGHTS

A. Concentration Camps

As a leader that keeps his people in line through a Machiavellian utilization of fear, Kim Jong-Il has seized a stranglehold on his raw power through concentration camps. The vast majority of inmates are considered political prisoners and are victims of what the UN defines as "arbitrary detention."[74] The concentration camps serve the purpose of thoroughly purging anyone standing against Kim Il-Sung and Kim Jong-Il's regime.[75] The plight of Hun-sik Kim, the principal of Pyongyang Light Engineering College, serves as an example. She was sentenced to five years imprisonment for simply suggesting to the City Education Board that her students' labor responsibility be reduced so that they could spend more time studying.[76]

Many inmates die before the completion of their sentence due to the horrendous living conditions inside the prison camps. The prison chamber is so congested that sleeping there is a torture in and of itself. "Some eighty to ninety prisoners sleep in a flea-infested chamber about six meters long by five meters wide (about 19 feet by 16 feet)."[77] Thus, the prisoners are forced to sleep with the stinking feet of other prisoners right under their nose. The prisoners naturally emit a stench due to the fact that they are only allowed to shower twice a year.[78] Additionally, prisoners are provided only enough food to leave them perpetually teetering on the verge of starvation.

---

[74] Hawk, *supra* note 14, at 26.

[75] *North Korea: Human Rights, Refugees, and Humanitarian Challenges: Joint Hearing Before the Subcomm. on Asia and the Pacific and the Subcomm. on Int'l Terrorism, Nonproliferation and Human Rights of the H. Comm. on Int'l Relations*, 108th Cong. 102 (2004) [hereinafter *North*], *available at* http://www.internationalrelations.house.gov/archives/108/93390.pdf.

[76] *North Korea: Humanitarian and Human Rights Concerns: Hearing Before the Subcomm. on East Asia and the Pacific of the H. Comm. on Int'l Relations*, 107th Cong. 62 (2002) [hereinafter *NKHR*], *available at* http://internationalrelations.house.gov/archives/107/79392.pdf.

[77] *Id.* at 60.

[78] *NKHR*, *supra* note 76, at 69.

Prisoners are compelled by their hunger to eat anything remotely edible, such as plants, grass, bark, rats, snakes, or food provided for the labor-camp farm animals, if they can get away with it.[79]

Not only are living arrangements horrendous inside these camps, but working conditions are frequently enough to kill inmates before the end of their sentence. The prisoners work around sixteen to eighteen hours every day without wages. Leather whips line the prison walls and inmates are beaten on a daily basis with little or no reason. The prisoners are forbidden to talk, laugh, or take a rest, and are required to keep their heads down at all times. These same motions are repeated throughout their workday. As a result, more than half of the women suffer from lumps on their heads or shoulders, they develop hunched backs, or they become crippled.[80]

Death by beating, starvation, and outright execution are common occurrences in the camps. Consider this report on Kyohwa-so Number 1, a labor camp in South Pyong-an Province:

> The whole group would be punished for the infraction of one of its members, a common infraction being the failure to meet individual—or group—production quotas. The most common and immediate punishment was reduced food rations. Frequently the threat of reduced food rations drove the women prisoners to work through constant pain. In winter, hands and fingers numb from cold were prone to accidents from the sewing needles and scissors. Mindful of their production quotas, prisoners continued at their work-stations, doubly fearful that their dripping blood would soil the garments they were sewing. Repeated infractions led to transfer to the prison's shoe factory. Even more severe punishment included prolonged solitary confinement in a cell too small to allow for a person to fully stand up or lie down inside, leading to loss of circulation and severe pain.[81]

---

[79] Hawk, *supra* note 14, at 25.

[80] *NKHR*, *supra* note 76, at 69.

[81] Hawk, *supra* note 14, at 44-45.

Pregnant women entering the concentration camps face particularly unfortunate conditions due to the unfathomably heartless treatment of newborns. For instance, concentration camp guards, nurses, and inmates are forced to murder all newborns and unborn babies who are believed to be of Chinese decent.[82] The following is a synopsis of the refugee testimony of a sixty-four-year-old grandmother:

> The first baby was born to a twenty-eight-year-old woman named Lim, who had been happily married to a Chinese man. The baby boy was born healthy and unusually large, owing to the mother's ability to eat well during pregnancy in China. Former Detainee #24 assisted in holding the baby's head during delivery and then cut the umbilical cord. But when she started to hold the baby and wrap him in a blanket, a guard grabbed the newborn by one leg and threw it in a large, plastic-lined box. A doctor explained that since North Korea was short on food, the country should not have to feed the children of foreign fathers. When the box was full of babies, Former Detainee #24 later learned, it was taken outside and buried.[83]
>
> She next helped deliver a baby to a woman named Kim, who also gave birth to a healthy full-term boy. As Former Detainee #24 caressed the baby, it tried to suckle her finger. The guard again came over and yelled at her to put the baby in the box. As she stood up, the guard slapped her, chipping her tooth. The third baby she delivered was premature—the size of an ear of corn—and the fourth baby was even smaller. She gently laid those babies in the box. The next day she delivered three more very premature babies and also put them in the box. The babies in the box gave her nightmares.[84]

B. The Plight of Refugees

Under the oppression of the feared demagogue, Kim Jong-Il, thousands of North Koreans flee their home country every year in search of a better life. The combination of extreme hunger, fear, and potential economic opportunity motivates refugees to risk their family's lives by making the dangerous trek to China, or to abandon their families and make their journey alone.[85] This unfortunate situation allows human traffickers the opportunity to exploit refugees, of whom nearly seventy-five percent are women. In fact, seventy to ninety percent of women who venture into China fall prey to sex trade human traffickers.[86]

---

[82] *Id.* at 61.

[83] *Id.* at 62.

[84] *Id.*

The Chinese government forcibly repatriates North Koreans seeking refuge in China, contravening its obligations under the 1951 Convention Related to the Status of Refugees and its 1967 Protocol. It is estimated that well over one-hundred fifty thousand North Korean refugees currently live in China. If repatriated, refugees face imprisonment, torture, or execution.[87] In December of 2003, the Commission to Help North Korean Refugees (CNKR) reported that over 850 North Korean refugees had been captured by Chinese security forces and were being held in five separate Chinese detention centers in the Yenbian region.[88] Well-informed sources reported that the refugees were being repatriated from the five camps to North Korea at a rate of roughly one hundred per week. Each week, fifty to one hundred additional refugees are reportedly repatriated from Dandong, China to Sinuiju, North Korea.[89]

Upon arrival in North Korea, refugees are subjected to intense interrogations that may last for several days or even weeks.[90] Torture is a staple of the interrogation process, along with sexual humiliation and exploitation of female detainees, as well as the horribly common practice of infanticide.[91] Reports indicate that the interrogators ask very similar questions such as: Why did you leave? Were you trying to get to South

---

[85] U.S. CONG. RESEARCH SERV., NORTH KOREAN REFUGEES IN CHINA AND HUMAN RIGHTS ISSUES: INTERNATIONAL RESPONSE AND U.S. POLICY OPTIONS (2007), *available at* http://www.fas.org/sgp/crs/row/RL34189.pdf.

[86] *North*, *supra* note 75, at 76.

[87] *Id.* at 2.

[88] *Id.* at 76.

[89] *Id.*

[90] Hawk, *supra* note 14, at 63.

[91] *Id.* at 59.

Korea? Did you talk to any South Koreans? Did you talk to any Christians? Did you listen to South Korean television or radio? Did you attend a Christian church?[92] If at any time the interrogators become convinced that the detainee has either been influenced by South Koreans or a community of faith, that individual will face either execution or banishment to one of the North Korean concentration camps.[93] Those that avoid this fate are sent to shorter-term work camps as the least cruel option.

C. Religious Persecution

Although religious freedom is guaranteed by the constitution, it is sharply curtailed in practice. Refugees and defectors held on the basis of their religious beliefs generally are treated worse than other inmates. Many Christians are reportedly being held in labor camps due to their religious beliefs.[94] A North Korean concentration camp prisoner explains:

> I have seen many scenes of Christians being punished because they would not change their belief. They would not say, okay, I will not believe in Christ anymore, and that is what the prison guards wanted to hear. I have seen eight women who were dragged out and being punished because they did not say or they did not say they would not believe in Christ anymore. These women were burned.[95]

Refugees who convert to Christianity during their fugitive life in China face a particularly grim fate when repatriated to their home country of North Korea. For instance, a family of four refugees practiced their religion without interference for over a year while under the protection of an undercover missionary in China until May 2002 when they were discovered and detained by Chinese police.[96] The family attempted to retain some of their religious reading by hiding it in their clothing; however, it was discovered by investigators from the North Korean State Security Agency.[97] Countless repatriated refugees have testified that the very first questions asked by interrogators are whether they had any contact with Christians while in China or if they believe in Jesus Christ.[98] Although many newly converted refugees hesitate, this family was firm and forthright in its profession of faith.[99] Following a bold declaration of faith in Jesus Christ, the authorities led the family to so-called "Hepatitis Street," a small courtyard adjacent to the liver ward of a hospital in Namyang City.[100] As the neighbors were forcibly assembled, gunshots rang out and all four family members fell to the ground with mortal wounds to the head.[101] "The message to the stunned cluster of neighbors was unmistakably clear: anyone who attempts to exercise a religious belief other than the worship of the Dear Leader, Kim Jong-Il, will meet the same fate."[102]

---

[92] *Id.* at 8.

[93] *Id.*

[94] *The Hidden Gulag: Putting Human Rights on the North Korea Policy Agenda: Hearing Before the Subcomm. on East Asian and Pacific Affairs, Comm. On Foreign Relations*, 108th Cong. 42 (2003) [Hereinafter *Gulag*], (statement of David Hawk, Human Rights Investigator, U.S. Comm. for Human Rights in North Korea), *available at* http://ftp.resource.org/gpo.gov/hearings/108s/92834.pdf.

[95] *Examining the Plight of Refugees: The Case of North Korea: Hearing Before the Subcomm. on Immigration, Comm. on the Judiciary*, 107th Cong. 28 (2002) (statement of Sun-ok Lee), *available at* http://www.access.gpo.gov/congress/senate/pdf/107hrg/86829.pdf.

D. Crimes Against Humanity

Massive and varied crimes against humanity are taking place in North Korea on a regular basis. The particularly inhumane and uncivilized examples are human extermination, torture, crimes by association, and collective retribution.[103] Without exception, crime-by-association is applied to all political crimes.[104] The concluding observations from the UNHRC's 2004 report demanded North Korea guarantee that all unfair treatments, tortures, and other inhuman treatments be swiftly reported to and investigated by an independent organization.[105] Nonetheless, North Korea fails to protect its citizens from the most severe human rights abuses.

---

[96] *North*, *supra* note 75, at 76-77.

[97] *Id.*

[98] *Id.*

[99] *Id.*

[100] *Id.*

[101] *Id.*

[102] *Id.*

[103] KINU 2004, *supra* note 9, at 9.

[104] *Id.*

Kim Jong-Il currently stewards at least one of, if not *the* most oppressive regimes in the world. In spite of North Korea's every attempt to hide this repulsive reality, human rights violations are numerous and well documented. In North Korea, due to guilt by association, children die of starvation in concentration camps because their grandfather or grandmother was deemed anti-Party. "People are executed in public because they ask the question: Why is the Great Leader not giving us rations?"[106]

1. Extermination

In 2001, it was estimated that the North Korean population shrunk by about seven million over the last ten years, and that at least two and a half million, or ten percent of the population, were sacrificed during that period.[107] North Korea is strongly rejecting the charge of public executions as absolute fabrications of hostile elements.[108] However, direct, firsthand evidence indicates that public executions are decided at summary trials, and most political prisoners know neither the nature of their crime nor the scope of their sentence. In this way, all suspects are put through inhumane treatment.[109] The following is the testimony of a North Korean political prisoner:

> In May 1998, three officials came to our camp for a test. These officials were not part of the State Security Bureau that is assigned to oversee the camps. They were dressed in light brown protective garb. The prison guards brought 50 women into an auditorium and then the guards left. The officials instructed me to give out to 50 women a liquid saturated cabbage that was in white buckets. When I handed this to the woman [sic] the officials told them to eat it. After the woman

---

[105] *Id.* at 96-97.

[106] *North*, *supra* note 75, at 100.

[107] KINU 2004, *supra* note 9, at 112.

[108] *Id.* at 91.

[109] *North, supra* note 75, at 3.

> [sic] ate it, blood came out from their mouths and their rectums. It looked like something had exploded inside them. In just a few minutes, they were all kneeling and falling forward. The blood that came out of them went for 5 to 6 feet. There was pandemonium and screaming. I was taken out of the auditorium and told to wash my hands. The next day the Camp Security Director and the three individuals were meeting and congratulating each other and saying that the results were greater than expected, the efficacy even greater than expected.[110]

The UN review subcommittee urged North Korea to revise the criminal code provisions relating to capital punishment in a way that would be consistent with the relevant articles on international covenants; however, North Korea has failed to do so.[111] Amnesty International has received reports of public executions carried out at places where large crowds gather. These executions are announced in advance to encourage attendance by schools, enterprises, and farms. Additionally, some prisoners have reportedly been executed in front of their families.[112] Executions are carried out by hanging or firing squad.[113] So many of these executions are for political, trivial, or manifestly unjust reasons and take place without the benefit of legal protections that would safeguard the innocent.

2. Torture

The North Korean government consistently argues that torture and human rights violations have disappeared as a result of relevant law revisions and the thorough education of law enforcement officers.[114] However, the Office of the United Nations High Commissioner for Human Rights has officially acknowledged the widespread human rights violations that regularly occur in North Korea. In fact, individual rights are considered subversive to the goals of the state and the party.

---

[110] *North*, *supra* note 75, at 111.

[111] KINU 2004, *supra* note 9, at 91.

[112] *Gulag*, supra note 94, at 42.

[113] *Id.*

[114] KINU 2004, *supra* note 9, at 96.

"The regime uses extreme repression and a pervasive surveillance network to intimidate and instill fear in the population. It maintains control through terror, threat of severe punishment, and the manipulation of privileges, including the privilege of food allotments."[115]

Torture and cruel punishment of political prisoners is an everyday occurrence in North Korean concentration camps.[116] Consider this report of a North Korean political prisoner:

> One day in early March 1997, I was taken into a torture chamber that I had never been in before. I saw a big kettle on a small table and a low wooden table with straps, about 20 centimeters high. By surprise, one of the two interrogators tripped me with his leg. They strapped me on to the table and forced the kettle spout into my mouth. The spout was made so that it forced my throat wide open and I could not control the water running into my body. Close to suffocation, I had to breathe through my nose. My mouth was full of water and it overflowed from my nose. As I began to faint from the pain and suffocation, I could not see anything but felt sort of afloat in the air. I had been through all kinds of torture, such as whippings, beatings with rubber bands or hard sticks, or hand twisting with wooden sticks between my ten fingers, but this was worse. I do not remember how long it lasted but when I woke up I felt two interrogators jumping on a board which was laid on my swollen stomach to force water back out of my body. I suddenly vomited and kept vomiting with terrible pain. I had no idea how much water ran into my body but I felt like the cells in my body were full of water and water was running out of my body through my mouth, nose, anus and vagina. I faintly heard somebody saying, 'Why doesn't this bitch wake up. Did she die?' I could not get up so I was dragged to my cell that day. From that day on, I suffered from high fever and often fainted. My whole body was so swollen that I could not open my eyes. I could only urinate a few drops of milk-like liquid with blood and felt a severe pain in my bladder. I was able to get up and walk again in about two week's time. I cannot explain how I could have survived such

---

[115] *NKHR*, *supra* note 76, at 4.

[116] *See generally* Hawk, *supra* note 14.

> an ordeal. I would have died if that had happened to me in my ordinary life. I must have developed a mysterious super power to sustain myself under an emergency situation.[117]

Many more first-hand accounts such as this one may be stated, but this one suffices as an emblematic example.

The following gives another instance of the egregious practice of torture by the North Korean government:

> In 1987, a school principal in Chongjin city found two female teachers murdered the previous night in the night duty chamber of the school. He immediately reported the murders to the police. When the police made little progress in the investigation, they arrested him for murder. He was subject to all kinds of severe torture for two years and forced into confessing the murder. When I saw him in the police jail, both his ears were gone with only ear holes in their place. I have no idea how it happened but his fingers were cut short and clustered together. He was badly crippled, one leg shorter than the other, and unable to walk. His mouth was slanted and he could not control his lips, which made it very difficult to understand what he said. He was a tall and handsome person before he was arrested but became as short as a ten year-old boy in the two years in the police jail.[118]

One common form of torture is to tie a prisoner against iron bars, spread his hands and legs apart, and beat him repeatedly all over the body with a rubber or cow skin whip. Just the pain from hanging by his body weight makes the ordeal unbearable. The beatings cause the prisoner's skin to tear and gush blood to the point where his skin does not even feel human anymore. When a prisoner is released from the iron bar, his whole body is so swollen that he cannot bend his back or knees. As a result, the prisoner must evacuate and urinate while standing.[119]

As a result of extended torture, North Korean citizens look incredibly stunted. People who are age nineteen or twenty look like they are twelve.[120] Many more examples may be adduced, but for present purposes, they will not receive further exposition.

---

[117] *NKHR, supra* note 76, at 65.

[118] *Id.* at 66.

[119] *Id.*

3. Crimes by Association and Collective Retribution

Citizens of North Korea face the fear of punishment for "crimes" committed by association or through collective contribution. That is, if the crime committed is serious enough, a relative of the criminal may also be punished[121] which may extend through three generations.[122] This is the most effective means to prevent complaints and collective actions and is applied to all political crimes.[123] Suicide is a crime that follows this method of punishment because, in North Korea, suicide is considered treason under the view that the act is regarded as a complaint by an individual about society and thus is a political crime. As such, those related to a person who commits suicide are subject to punishment for their family member's crime.[124] Additionally, to say a person died of starvation is considered a political crime so it is never stated as the cause of death by the deceased's family members out of fear they will be punished.[125]

IV. CONCLUSION

When friends asked about this work on human rights in North Korea, I half-facetiously quipped that the essay would be one word long: "none." This grim attempt at humor aside, I did consider the state of North Korea to merit another word that made its way into the title of this piece, this time a somewhat synonymous neologism: "rightlessness." The term "state" preceding rightlessness contains an intentional *double entendre*: it indicates both "a situation of" as well as "a government." The adjective "egregious," also in the title, sadly does not qualify as hyperbole but a proper description, as this essay has succinctly sketched.

---

[120] *Id.* at 10.

[121] KINU 2004, *supra* note 9, at 11.

[122] Windrem, *supra* note 19.

[123] KINU 2004, *supra* note 9, at 11.

[124] *Id.*

[125] *Id.* at 12.

This writing serves as a prelude to a companion work to follow. The sequel offers a possible means to address the state of rightlessness, the egregious case of North Korea: a hybrid tribunal that would ideally combine the strengths of a domestic court with the contributions of an international court to enhance legitimacy, build domestic capacity, reduce costs, and help bring a measure of justice. A court of this type must carefully contextualize itself to the particulars of Korea and avoid the pitfalls of imperialism, corruption, and strong bias.

Such a hybrid tribunal and the aspiration to strengthen Korea's domestic legal system can also contribute momentum towards the eventual formation of an East Asian system of human rights. ASEAN has taken very recent, small steps towards the genesis of a system for Southeast Asia. Drawing from the lessons of already existing regional systems of human rights in Europe, the Americas, and Africa, such a system for East Asia could serve as a more permanent bulwark to fortify respect for human rights in the region. It could in time reduce the need for, and perhaps someday obviate the sorts of hybrid tribunals mentioned herein.

Until the establishment of such a system of human rights for the most populous continent in the world though, a means like a hybrid tribunal can serve as a stopgap measure to address the hemorrhaging in places such as North Korea. The unspeakable suffering of so many under Pyongyang's steel boot should not be met with passive indifference. Whether through the International Criminal Court (ICC), domestic prosecution, or a hybrid tribunal, impunity and oppression should not be allowed to reign unchecked. The healing of the Land of the Morning Calm, the stability of the region, and the sizable ramifications around our deeply interconnected world all hang in the balance. If a threat to justice anywhere is a threat to justice everywhere (with due credit to Martin Luther King, Jr.), this state of rightlessness, this egregious case of gross, systematic violations of human rights in North Korea, simply demands redress. That redress serves as the subject for the law review article that follows this one.

COMMENT

TRANSMUTING MISSISSIPPI'S CURRENT TRANSMUTATION DOCTRINES: ESTABLISHING CLEAR AND CONSISTENT PRECEDENTS TO PROPERTY DIVISION

> When . . . a question has once been judicially considered and answered, it must be answered in the same way in all subsequent cases in which the same question again arises. Only through this rule can that consistency of judicial decision be obtained, which is essential to the proper administration of justice.[1]

I. INTRODUCTION

Imagine a married couple—John and Judy—who have been married for ten years and who moved from Louisiana to Mississippi approximately five years ago. John works for a car company in Jackson, and Judy is a stay-at-home mother of two. About three years ago, Judy inherited from her deceased mother an expensive painting, a two-story house worth $200,000 encumbered with a $10,000 mortgage, and $100,000 in cash. The couple promptly moved into the home after the completion of probate, and Judy placed the expensive painting above the mantle as a stunning showpiece for the couple's house. Judy also deposited the $100,000 into the couple's joint account for three months but then removed the funds, placing the $100,000 into CDs under her name only. Unfortunately, the couple's relationship deteriorated a year ago, and John filed for divorce based on irreconcilable differences. The question raised is whether the expensive painting, the two-story home, and the

---

[1] JOHN W. SALMOND, JURISPRUDENCE OR THE THEORY OF THE LAW 171 (Steven & Haynes, eds., 1907).

$100,000 in CDs are separate property or marital property? Regrettably, the answer is not crystal-clear due to Mississippi's current transmutation laws.[2] Mississippi's transmutation laws are inconsistent, making it difficult if not impossible to predict the eventual outcome of the case at both the trial and appellate levels.

This lack of consistency has several undesirable effects. First, it leads to uncertainty for legal practitioners and their clients, and arbitrary decisions by lower-level chancellors.[3] Secondly, the lack of conformity increases litigation rather than fostering settlement, and leads to undesirable social effects such as married couples not using valuable separate property assets for familial use in an attempt to preserve its separate property status.[4] Finally, Mississippi's transmutation laws are not aligned with divorcing parties' expectations of property division, thus resulting in unwanted and unnecessary confusion for both parties.[5]

Part II of this article examines Mississippi's current transmutation laws—both the doctrines of commingling and the family-use doctrine. Part III examines arguments advanced by prior legal commentators about Mississippi's transmutation laws as they exist today. Part IV offers an analysis of the transmutation laws by way of illustration to our hypothetical couple—John and Judy. Part V advocates for change to Mississippi's transmutation doctrines and demonstrates various pitfalls of the law as it exists today. Part VI offers a comparison of Mississippi laws with other equitable distribution states. Finally, Part VII concludes by recommending change through abolishing the family-use doctrine and the unitary approach to commingling, and advocating for clear and consistent precedent during the property division stage.

---

[2] Transmutation is defined as "[a] change in the nature of something; esp., in family law, the transformation of separate property into marital [or vice versa]." BLACK'S LAW DICTIONARY 1537 (8th ed. 2004). Transmutation in Mississippi can occur through either the family-use doctrine, discussed *infra* Part II.B., or through commingling of separate property assets with marital property assets, also discussed *infra* Part II.A.

[3] *See* 3 BRETT R. TURNER, EQUITABLE DISTRIBUTION OF PROPERTY app. A at 255 (3d ed. 2005). Mississippi trial courts "may recognize commingled separate property if and when [they] want[ ] to." *Id.*

[4] *Id.* Because the property system adopted by Mississippi is "*so inconsistent, it makes divorce cases much harder to settle.*" *Id.* (emphasis in original). Turner continues that parties "who use their separate property freely are greatly penalized" and thus "generous and productive behavior is penalized, while selfish behavior is rewarded." *Id.*

[5] *See infra* Part V.A.

II. CURRENT TRANSMUTATION LAW IN MISSISSIPPI

In 1994, the Mississippi Supreme Court significantly altered the state's family law jurisprudence—adopting equitable distribution as the sole method of dividing marital assets after divorce.[6] With this result came the instruction that courts are to classify each asset as either marital or separate, with a strong but rebuttable presumption that all assets are marital.[7] Assets derived during the marriage are presumed marital properly unless one party can demonstrate that the assets were obtained through that party's separate estate "prior to the marriage or outside the marriage."[8] The distinction between marital and separate property is significant because only marital assets are subject to property division.[9] Separate property assets include inter vivos gifts to one spouse;[10] inheritance and gifts bestowed on one individual party;[11] separate property exchanged for other property;[12] or separate property accumulated before the marriage.[13] Nevertheless, separate property assets do not necessarily retain their status as separate indefinitely, but can become part of the marital estate through a process called transmutation. Transmutation of property in Mississippi can occur in one of two ways—the family-use doctrine or commingling.[14] The family-use doctrine converts separate property into marital when family members use separate property extensively for familial purposes.[15] Additionally, separate property may lose its status when that property is commingled,[16] or mixed, with marital property.[17]

---

[6] *See* Ferguson v. Ferguson, 639 So. 2d 921, 925 (Miss. 1994) ("With this opinion, [the] Court adopts guidelines for application of the equitable distribution method of division of marital assets."). Additionally, the Court issued guidelines in an effort to guide "chancellors in their adjudication of marital property division." *Id.* at 928. "Under the equitable distribution system the marriage is viewed as a partnership with both spouses contributing to the marital estate in the manner which they chose." *Id.* at 927 n.4.

[7] Hankins v. Hankins, 866 So. 2d 508, 511 (Miss. Ct. App. 2004) ("[T]he first duty of the chancellor is to categorize each asset as either marital or non-marital based upon the evidence presented."). "The burden [of proof is on the party] claiming [the] assets to be non-marital to demonstrate . . . their non-marital character." A & L, Inc. v. Grantham, 747 So. 2d 832, 839 (Miss. 1999). *See also* DEBORAH H. BELL, BELL ON MISSISSIPPI FAMILY LAW § 6.02, at 155 (1st ed. 2005).

[8] Parker v. Parker, 980 So. 2d 323, 328 (Miss. Ct. App. 2008); Wideman v. Wideman, 909 So. 2d 140, 144 (Miss. Ct. App. 2005). *See also* Johnson v. Johnson, 650 So. 2d 1281, 1286 n.2 (Miss. 1994) ("[P]roperty clearly obtained by one party through inheritance or acquired by one party by gift is nonmarital property not subject to equitable distribution.").

[9] Messer v. Messer, 850 So. 2d 161, 167 (Miss. Ct. App. 2003).

---

[10] *See Johnson*, 650 So. 2d at 1286 (holding husband's marital monetary property gift to wife separate property).

[11] *See* Everett v. Everett, 919 So. 2d 242, 248 (Miss. Ct. App. 2005) (holding wife's inherited house and other assets separate property because they derived from inheritance); Baldwin v. Baldwin, 788 So. 2d 800, 805-06 (Miss. Ct. App. 2001) (reversing trial court's determination that property gifted to husband by his mother was marital property).

[12] *See* Carrow v. Carrow, 642 So. 2d 901, 907 (Miss. 1994) (explaining property originally acquired through separate property funds treated as separate property).

[13] *See* Hankins v. Hankins, 866 So. 2d 508, 511 (Miss. Ct. App. 2004) (holding husband's premarital owned chicken farm retained separate property classification).

[14] *See* BELL, *supra* note 7, § 6.04[1][b], at 163-64, § 6.04[3], at 165.

[15] *See, e.g.*, Heigle v. Heigle, 654 So. 2d 895, 897 (Miss. 1995) ("[N]onmarital assets . . . may be converted to marital assets if they are . . . used for familial purposes."); *Johnson*, 650 So. 2d at 1286 (property brought into the marriage by one partner that is separate but used by the family may become marital); *see also* Pittman v. Pittman, 791 So. 2d 857, 866-67 (Miss. Ct. App. 2001) (finding separate property furniture placed in marital home was marital property absent evidence showing spouse who owned separate property used it exclusively). The court in *Pittman* also held if separate property china and silver were stored, it would probably keep its separate property status, but if used for entertaining or on other occasions, "issues of its becoming marital property arise." *Id.* at 867.

[16] Commingled property is defined as "a combination of marital and non-marital property in which the non-marital property has lost its status as non-marital property by virtue of its being combined with the marital property." Boutwell v. Boutwell, 829 So. 2d 1216, 1221 (Miss. 2002).

[17] *Id.* ("[Transmutation may occur when non-marital assets are] commingled with marital property . . . absent an agreement to the contrary.").

A. Transmutation Through Commingling

Not long after the landmark *Ferguson* decision, the Mississippi Supreme Court held in *Johnson v. Johnson* that separate property commingled with marital property becomes susceptible to equitable distribution.[18] Later cases have shown the commingling doctrine to apply when one party used marital funds to pay property taxes on separate property real estate;[19] when a couple contributed marital funds to assist paying off mortgage payments on a wife's separate property residence and the husband made improvements to the residence;[20] and when a spouse used money gifted to her by her parents to pay for the marital home.[21] The supreme court and court of appeals have continued to follow the established precedent of *Johnson* and its progeny to some extent, but notable exceptions, along with ambiguity, have arisen.

One such notable exception occurred in *Bresnahan v. Bresnahan*.[22] In *Bresnahan*, the husband had placed rental income from two separate property homes into a marital bank account—an account that received income from marital property rental homes and also paid for improvements on both the marital and separate property rental homes.[23] At trial, his wife argued the holding of *Johnson* was applicable and argued that the separate property rental homes had been transmuted into marital property.[24] The supreme court disagreed, holding that the "co-mingling of rent from the various rental properties [did] not rise to the level of commingling in [prior precedent]."[25]

---

[18] *Johnson*, 650 So. 2d at 1286. In *Johnson*, Jane Johnson had inherited $153,000 from timber sales which she used for familial expenses and also placed part of the funds in savings accounts. The court concluded that when Jane "commingled the monies received from the sale of timber, she converted these nonmarital assets to marital assets, subject to equitable distribution." *Id.* The court also found when Wayne Johnson gave $66,000 back to Jane, the property "reacquired its nonmarital character by virtue of his actions." *Id.*

[19] *See* Messer v. Messer, 850 So. 2d 161, 168-69 (Miss. Ct. App. 2003).

[20] *See* Belding v. Belding, 736 So. 2d 425, 432 (Miss. Ct. App. 1999).

[21] *See* Henderson v. Henderson, 703 So. 2d 262, 265 (Miss. 1997).

[22] Bresnahan v. Bresnahan, 818 So. 2d 1113 (Miss. 2002).

[23] *Id.* at 1117-18.

[24] *Id.* For a discussion of *Johnson*, *see supra* notes 15 & 18.

[25] *Id.* at 1118.

The same year *Bresnahan* was decided, the court of appeals put a further caveat on the commingling doctrine. In *Oliver v. Oliver*, the Court of Appeals held:

> [w]e are not convinced that the Mississippi Supreme Court meant that any funds obtained by one spouse, which are subsequently commingled with marital funds, automatically become marital property simply because they are placed in the same account. We opine . . . that a party must prove that these . . . commingled funds were not only present in a joint marital account, but were being used for the benefit of the other spouse and/or the entire family.[26]

Oliver is unique in that the court interpreted and clarified family law doctrines in a succession dispute case. Helen Oliver had inherited approximately $70,000 from a close friend and subsequently placed the funds in a checking account that belonged to her, her husband B.A., and her husband's son, Roger.[27] Helen left these funds in the joint bank account for a short period of time and then converted the funds into certificates of deposit in the name of her husband, Roger, and herself.[28] Eventually, Helen decided to use these funds to purchase an asset account with Occidental Life Insurance Company.[29] Thereafter, Helen withdrew the money and terminated the policy.[30] B.A. died intestate soon after these transactions occurred, and Roger argued that through the doctrine of commingling, he was entitled to fifty percent of the funds Helen withdrew from the Occidental Life Insurance account.[31] The court of appeals held that the money inherited by Helen and placed in the bank account for a short period of time was never intended for B.A. or

---

[26] Oliver v. Oliver, 812 So. 2d 1128, 1135 (Miss. Ct. App. 2002).

[27] *Id.* at 1131.

[28] *Id.*

[29] *Id.*

[30] *Id.*

[31] *Id.* at 1131-32. Roger argued that the funds inherited by Helen were commingled when Helen placed the funds in B.A. and Helen's marital bank account. *Id.* at 1132.

Roger's benefit, and thus was never a part of B.A.'s estate.[32] As such, Roger had no claim to the funds.[33]

Another recent decision by the court of appeals further blurred the distinction between separate property and marital property with its holding in *Brock v. Brock*.[34] In *Brock*, the husband, J.D., argued that the chancellor erred when he refused to order any property interest in the couple's marital home to J.D.[35] The marital home was owned by Robin, J.D's wife, and had been deeded to Robin by her father while the couple was married.[36] J.D. argued that when he made repairs and paid taxes on the home, the home lost its separate property classification and was transmuted into marital property.[37] The court of appeals, surprisingly, held that "the key to determining when there has been transmutation by commingling is whether the marital interests can be identified, i.e., can be traced."[38] The court reasoned that J.D. had made "minimal" repairs to the residence and that his "contributions to the home by paying property taxes [were] readily traceable."[39]

As seen through Mississippi jurisprudence, current Mississippi law adheres to the broad policy of categorizing commingled property as entirely marital but with numerous ambiguous exceptions chancellors may or may not apply.[40] As Brett

---

[32] *Id.* at 1134.

[33] *Id.*

[34] Brock v. Brock, 906 So. 2d 879 (Miss. Ct. App. 2005). The decision of *Brock v. Brock* is revisited again later in this comment due to the court of appeals' interpretation of the family-use doctrine. *See infra* notes 50-52 and accompanying text.

[35] *Id.* at 887.

[36] *Id.* The home was also placed in Robin's name alone on the deed. *Id.*

[37] *Id.* at 888.

[38] *Id.*

[39] *Id.* The court cited *Brock* and referred to tracing again in *Dorsey v. Dorsey* when they held that a spouse's possible payment of property taxes towards a separate property residence would be "both minimal and easily traceable." Dorsey v. Dorsey, 972 So. 2d 48, 52-53 (Miss. Ct. App. 2008).

[40] *See* 3 TURNER, *supra* note 3, app. A at 253 ("It is now time to recognize, formally and expressly, that under the law of Mississippi, all mixtures of marital and separate property become entire [sic] marital."); *see also* BELL, supra note 7, § 6.04[3], at 165. Mississippi follows the rule, with some exceptions, that when separate property touches marital property, it becomes marital property, regardless of whether "the separate property is identifiable." *Id.*

Turner mentions, the *Brock* decision is particularly puzzling because countless other Mississippi court opinions have failed to "pay the slightest attention to traceability of the separate contribution" whereas others have based a large part of their reliance on the tracing doctrine.[41] Disregarding the various exceptions Mississippi cases have provided, Mississippi's unitary commingling treatment of property uniquely places it in an overwhelming minority compared to other states across the United States.[42]

B. Transmutation Through the Family-Use Doctrine[43]

In 1994, the supreme court recognized the family-use doctrine, to a limited extent, in *Johnson v. Johnson.*[44] A year later the supreme court affirmed and broadened its holding in *Johnson* when it stated "nonmarital assets . . . may be converted to marital assets if they are . . . used for familial purposes [and as a result] [s]uch converted assets are . . . subject to equitable distribution."[45] Since 1995, the supreme court and court of appeals have found the family-use doctrine applicable when a spouse inherited a separate property residence but subsequently both spouses worked on the home and resided in the home,[46] and when separate property antique furniture was placed in the marital home and used by the family.[47] Additionally, the court of appeals recognized in dicta that the family-use doctrine could apply to separate property real estate that the couple's children rode over on four wheelers.[48] Noteworthy, however, are the numerous appellate cases that have diverged from this all-or-nothing approach[49] without specifically overruling prior precedent.

---

[41] 3 BRETT R. TURNER, EQUITABLE DISTRIBUTION OF PROPERTY app. A at 78 (Supp. 2009-2010). *Brock* is not the sole Mississippi opinion citing the tracing doctrine to divide property. *See supra* note 39.

[42] BELL, *supra* note 7, § 6.04[3][b], at 166-67. In a majority of states, "separate property becomes marital when it is so commingled with marital property that the owner cannot trace the separate funds." *Id.* § 6.04[2][3], at 165. If the separate property "can be identified, the asset is classified as mixed, partly marital and partly separate." *Id.* § 6.04[3][a], at 166.

[43] "Mississippi courts have created a unique conversion rule by focusing on a single element of the implied gift theory—family use." BELL, *supra* note 7, § 6.04[b], at 163-64.

[44] *See* Johnson v. Johnson, 650 So. 2d 1281, 1286 (Miss. 1994). The supreme court declared that timber sales from separate property inherited by one spouse were transmuted into marital property when the money was used "for the benefit of the entire family." *Id.*

[45] Heigle v. Heigle, 654 So. 2d 895, 897 (Miss. 1995).

[46] *See* Boutwell v. Boutwell, 829 So. 2d 1216, 1221 (Miss. 2002); *see also* McDuffie v. McDuffie, 21 So. 3d 685, 691 (Miss. Ct. App. 2009) (separate property home brought into marriage by party was converted to marital property through use by family). The court of appeals has recognized that the "family-use doctrine will almost always con-

For example, in *Brock v. Brock*, the court of appeals, in interpreting the family-use doctrine, stated that the "marital home is not necessarily marital property."[50] In *Brock*, the husband, J.D., appealed a decision by the lower court that awarded his wife, Robin, full ownership of the marital home.[51] J.D. argued during appeal that he was entitled to an interest in the home because of the family-use doctrine but the court affirmed the decision of the trial court, reasoning that "[t]he house was intended to be a gift to Robin alone," despite the fact that the couple had occupied the home as a family for a period of years.[52]

Adding only to the ambiguity of the family-use doctrine is the result of *Ory v. Ory*.[53] Sharon Ory owned seventy-five acres

---

vert a separately owned 'marital' home to marital property." Faerber v. Faerber, 13 So. 3d 853, 861 (Miss. Ct. App. 2009) (internal citations omitted)). The commingling doctrine also increases the chances a separate property home will become marital property. This results because payments for the mortgage, property taxes, and upkeep of the residence will most likely be from marital funds. Thus, both doctrines usually apply to marital homes.

[47] *See* Pittman v. Pittman, 791 So. 2d 857, 866 (Miss. Ct. App. 2001).

[48] *See* Dorsey v. Dorsey, 972 So. 2d 48, 52 (Miss. Ct. App. 2008). The court declined to reverse the decision of the lower court holding the property separate because there was no "corroborating testimony establishing the frequency and extent to which the children used the land in question." *Id.*

[49] In a concurring opinion, Judge Southwick labeled the family-use doctrine as "an all one or all the other approach." Hankins v. Hankins, 866 So. 2d 508, 513 (Miss. Ct. App. 2004) (Southwick, P.J., concurring). Judge Southwick's concurring opinion is discussed extensively *infra* at Part III.

[50] Brock v. Brock, 906 So. 2d 879, 887 (Miss. Ct. App. 2005). *But see Boutwell*, 829 So. 2d at 1221 (holding that a residence inherited by the wife but occupied by both parties changed the character of the separate property home into a marital asset).

[51] *Brock*, 906 So. 2d at 887.

[52] *Id.*

[53] Ory v. Ory, 936 So. 2d 405 (Miss. Ct. App. 2006).

of property that had been in her family for more than 100 years.[54] Alan Ory argued transmutation applied due to various jobs he had performed on the land, but the chancellor classified the land as separate property despite evidence that demonstrated "Alan [had] cleared a portion of the land, hauled dirt onto the property, and had a large number of seedlings planted on the property."[55] The court of appeals held the chancellor was not in error "characterizing the parcel as a non-marital asset."[56] In coming to their conclusion, they stated "Alan did not . . . put forth sufficient evidence to prove that his activity was so pervasive as to convert the entire seventy-five-acre parcel into a marital asset."[57]

In addition, there have been recent decisions by the court of appeals that have not discussed the family-use doctrine or its application even though it readily appears it may have applied. For example, in *Duncan v. Duncan*, Kim Duncan had purchased a home prior to her marriage to Maurice Duncan, and Maurice and Kim resided in the home as husband and wife.[58] The parties filed for divorce, and Kim presented evidence at trial that demonstrated she had paid primarily all of the mortgage payments and renovations through her bank account.[59] The chancellor ruled that the home was to be awarded to Kim, and that Maurice was not entitled to any equitable interest in the home.[60] Surprisingly, the court of appeals, focusing on the commingling theory, affirmed the chancellor's decision, without any discussion of the family-use doctrine.[61]

---

[54] *Id.* at 411.

[55] *Id.* Alan also argued unsuccessfully that the commingling doctrine applied because of payments he made to settle a lien on the property. *Id.* Both the lower-court and court of appeals found this argument without merit. *Id.*

[56] *Id.*

[57] *Id.*

[58] Duncan v. Duncan, 915 So. 2d 1124 (Miss. Ct. App. 2005).

[59] *Id.* at 1126. Kim testified at trial that she had paid $12,000 in renovations whereas Maurice had paid $1000. *Id.* at 1125. Maurice testified that he had paid half the mortgage while the couple was married but was unable to provide proof by way of receipts. *Id.* at 1126.

[60] *Id.* at 1126.

[61] *Id.*

Finally, at issue in *Gutierrez v. Bucci* was the proper classification—separate or marital—of two Harley Davidson motorcycles.[62] The motorcycles had been purchased by Dr. Bucci with retirement funds acquired entirely before he was married to Dr. Gutierrez.[63] The chancellor classified both motorcycles as separate property, and Dr. Gutierrez argued on appeal that because she rode the two motorcycles "on occasion," transmutation had changed the character of the motorcycles to marital property.[64] The court of appeals disagreed and found that the "occasional riding of one of the motorcycles by the nonpurchasing spouse [does not] necessarily cause[ ] the separate property to lose that classification."[65]

Thus, current Mississippi jurisprudence dictates separate property may be converted into marital property under the family-use doctrine depending on the particular circumstances of the case. Critical factors chancellors inquire into are the degree of use and the length of time the separate property assets were used for family purposes—a decision that often times will vary according to the chancellor assigned to that particular case. As with Mississippi's comingling doctrine, Mississippi's application of the family-use doctrine places it in "a distinct minority."[66]

III. CHALLENGES TO CURRENT MISSISSIPPI TRANSMUTATION PRECEDENT

The apparent ambiguity and difficulty in classification of marital assets has not gone without its share of criticism. For example, in 2004, Presiding Judge Southwick authored a concurring opinion in *Hankins v. Hankins*—joined by four other justices—that questioned the reasoning behind Mississippi's transmutation doctrines.[67] In his opinion, Judge Southwick stated the family-use doctrine was "too aggressive a view of marital property rules" and that "[t]he use of [a] residence and its contents for marital purposes should not automatically cause either the residence or its contents to become totally marital absent other circumstances."[68] Southwick elaborated that "an all one or all the other approach" remained difficult to apply and that "[m]erely because a house is the residence of the marriage, . . . it should not be considered as totally marital property."[69] Southwick further advocated the possibility of treating separately owned property used by both parties as "property [that] takes on a dual character."[70]

---

[62] Gutierrez v. Bucci, 827 So. 2d 27, 37 (Miss. Ct. App. 2002).

[63] *Id.*

[64] *Id.*

[65] *Id.*

[66] *See* BELL, *supra* note 7, § 6.04[1][b], at 164.

[67] *See* Hankins v. Hankins, 866 So. 2d 508, 513 (Miss. Ct. App. 2004) (Southwick, P.J., concurring).

Southwick and the other concurring justices are not alone in advocating a change to Mississippi transmutation doctrines. Brett Turner, a notable family law scholar, cites several important policy concerns with Mississippi's unitary theory of property.[71] First, Turner argues that under the current system, trial judges are given too much discretion and therefore their decisions can be entirely arbitrary.[72] Turner reasons that "[s]ome trial judges will treat contributions of separate property as a very important division factor; some will threat [sic] them as a modest division factor; still others will generally ignore them."[73] The end result is "rampant inconsistency in the results of the cases."[74] Turner continues his frustration by stating the "law of the state functionally becomes: the trial court may recognize commingled separate property if and when it wants to" and "personal beliefs of the trial court judge are often a more important factor than the substantive content of the law."[75]

---

[68] *Id.*

[69] *Id.*

[70] *Id.* Brett Turner provides that the "'dual character' approach suggested by the concurrence is the mixed theory of property, the consensus rule in all other dual-classification equitable distribution states." 3 TURNER, *supra* note 3, app. A at 257 (citation omitted). Under a dual property approach, property retains both a marital portion and a separate property portion. BELL, *supra* note 7, § 6.02[4][b], at 158.

[71] *See* 3 TURNER, *supra* note 3, app. A at 253. Commentators refer to states that "treat all property as either entirely marital or entirely separate . . . [as a] *unitary theory of property*." *Id.* (emphasis in original).

[72] *Id.* at 255, 257.

[73] *Id.* at 255.

[74] *Id.*

Turner's criticism does not stop here, however, as he alludes to the "poor economic and social policy" of current transmutation laws in Mississippi.[76] In support of this theory, Turner cites the conundrum spouses with separate property face. On the one hand, they may "passively store their separate property," thus ensuring they "receive that separate property upon divorce."[77] However, spouses who choose to "use their separate property freely are greatly penalized; separate assets mixed with marital property become marital, even if the marital component of the mixture is comparatively small."[78] Thus, "generous and productive behavior is penalized, while selfish behavior is rewarded."[79] Under current Mississippi law, "owners of separate property have a very strong incentive to hoard that property in unproductive isolation, and not to use that property for family benefit."[80]

Finally, Turner points to the inconsistency of Mississippi's case law, and asserts "it makes divorce cases *much* harder to settle."[81] The only way for parties and attorneys "to find out how separate property [will] be treated [is] to try the case."[82] Turner concludes that other states have adopted unitary property systems for property division during divorce but all ended in failure and adoption of dual classification.[83]

---

[75] *Id.*

[76] *Id.* Turner also states that the "great majority of states have rejected the sort of family use doctrine which the Mississippi cases are adopting." 3 TURNER, *supra* note 41, app. A at 80.

[77] 3 TURNER, *supra* note 3, app. A. at 255.

[78] *Id.*

[79] *Id.*

[80] 3 TURNER, *supra* note 41, app. A at 80. *See also* Pittman v. Pittman, 791 So. 2d 857, 867 (Miss. Ct. App. 2001) (stating in dicta if separate property is stored away or not used that property retains its separate property classification).

[81] 3 TURNER, *supra* note 3, app. A at 255 (emphasis in original).

[82] *Id.*

[83] 3 TURNER, *supra* note 3, app. A at 254. Both Illinois and Virginia attempted to create a unitary property system for property division in divorce cases. *Id.* However, "[u]nder the unitary theory of property, the concept of separate property proved to be remarkably hollow." *Id.* (emphasis in original). As a result, "[i]n both states, unitary property was a complete and demonstrated failure, and it was reversed by the legisla-

IV. ILLUSTRATIVE EXAMPLES: A RETURN TO THE STORY OF JOHN AND JUDY

The article now returns to the question presented in the introductory paragraph, more specifically, is the expensive painting, the two-story house, and the $100,000 separate property or marital property?[84] As mentioned previously, the answer is not entirely clear, but the arguments advanced by both John and Judy's counsel will be predictable and warrant closer attention.

First, with regard to the expensive painting, John and his attorney will argue the painting underwent transmutation as a result of the family-use doctrine.[85] Specifically, they will argue the painting was placed in the marital home, and the family used the painting by admiring it above the mantle. To support this argument, John will cite cases such as *Pittman v. Pittman*, which held separate property antiques and furniture can become part of the marital estate due to family use of the property.[86] Judy and her counsel, on the other hand, will argue the painting was not used the requisite amount of time to qualify for marital property and thus has not undergone transmutation.[87] In support of her position, she will cite cases such as *Gutierrez v. Bucci* that held property only used "on occasion" does not automatically become marital property.[88] Thus, the chancellor will consider the degree and length of time the painting was used and which precedent he finds more persuasive.

Secondly, John will argue both commingling and the family-use doctrine apply to the marital home.[89] In support of this contention, he will argue marital funds were used to pay off the

---

ture within three years of its adoption." *Id.* The "opposition to unitary property [in both states] was led by family law practitioners, upset partly by the policy behind [the] rule but even more by the fact that [the] rule was simply not working in practice." *Id.* Illinois and Virginia's failure and subsequent repeal of unitary property is discussed further in Part VI.

[84] *See supra* Part I.

[85] *See supra* notes 43-66 and accompanying text.

[86] *See supra* note 47 and accompanying text.

[87] *See supra* notes 53-65 and accompanying text.

[88] *See supra* notes 62-65 and accompanying text.

[89] *See supra* notes 18-21, 43-48 and accompanying text.

remaining $10,000 mortgage that encumbered the property, and that marital funds were used to contribute to routine repair and maintenance. Thus, John will argue marital funds have been commingled with separate property, and that the separate property home should properly be classified as part of the marital estate. John will cite cases such as *Belding v. Belding* that have held marital funds used to pay off mortgage payments can change the character of separate property to marital.[90] John will also argue the family-use doctrine applies because John and Judy, as well as their two children, have resided in the home for a period of three years.

Judy will rebut John's arguments by providing that the court of appeals has recognized both parties can live in a marital home and that the family-use doctrine does not necessarily apply.[91] She will further argue that the funds John placed into the marital home can be traced and thus have not been commingled.[92] Once again, the issue comes down to which precedent the chancellor finds the most authoritative.

Finally, John will argue the $100,000 was commingled with marital property when Judy placed the funds into the marital bank account.[93] Judy will argue that the court of appeals has recognized that tracing can be applied to such a situation, especially when the funds were only placed in the marital account for a short duration of time.[94] Not surprisingly, the chancellor will determine the issue by which precedent she finds more persuasive.

Thus, all three issues will hinge upon the specific findings of the chancellor assigned to the case. Neither party will conclusively know to any reasonable degree of certainty how the property will be classified until the matter is tried, and the decision is rendered.

---

[90] *See supra* note 20 and accompanying text.

[91] *See supra* notes 50-61 and accompanying text.

[92] *See supra* notes 34-39 and accompanying text.

[93] *See supra* notes 18-21 and accompanying text.

[94] *See supra* notes 26-33 and accompanying text.

V. TRANSMUTING MISSISSIPPI DOCTRINES INTO CLEAR AND CONSISTENT PRECEDENT AND REASONS FOR CHANGE

Now that the reader has been introduced first-hand to the complexity of Mississippi's transmutation laws through use of examples, the next section of this article will examine compelling and persuasive reasons why Mississippi's transmutation laws should be altered in favor of clear and consistent precedent.

Achievement of clear and consistent precedent can only occur through revocation of the family-use doctrine and adoption of a dual character property system instead of unitary property. As seen through examples, the family-use doctrine in itself is a doctrine that lacks clarity and consistency. Chancellors are called upon to ascertain the degree of use and length of time it requires to convert an asset from separate property to marital property—factors that inherently require subjective determinations. Not surprisingly, the conclusion of each chancellor is different according to that chancellor's personal views and interpretation of the law.

Similarly, Mississippi's current classification of assets as all marital or all separate irrespective of the amount of separate property creates unnecessary confusion. This confusion has been exacerbated by the numerous exceptions chancellor's may or may not apply. Adopting a mixed property system during the property division stage creates stronger and consistent precedent that members of the judiciary and divorced parties can rely upon. The clear and consistent precedent has added benefits explored below which include: alignment with each parties' expectations, beneficial state social policies are enhanced, and promotion of parties settling their disputes.[95]

---

[95] *See, e.g.*, Larry Alexander & Emily Sherwin, *Judges as Rule Makers*, *in* COMMON LAW THEORY 33-34 (Douglas E. Edlin ed., 2007). The authors discuss several potential benefits of consistent precedent, labeled by the authors as "entrenched rules." "Entrenched rules can reduce the cost of judicial decision making and, at least in some circumstances, encourage private settlement." *Id.* at 33. "By limiting discretion, precedent rules also reduce the opportunity for bias in decision making and correspondingly increase public faith in the impartiality of courts." *Id.* at 33-34.

A. Parties' Expectations

Divorce laws—particularly those dealing with property division—should strive for both fairness and alignment with each party's expectations.[96] As Robert Levy proclaims,

> It may not be universal, but it is certainly not uncommon for couples to view income that either of them produced during the marriage as 'of the marriage' and different from funds or assets brought by one of the spouses to the marriage or acquired by one of the spouses during the marriage by gift or inheritance.[97]

Returning to our example with Judy, it is a reasonable assumption that Judy failed to consult an attorney or discover Mississippi family law after moving to Mississippi and thus lacks knowledge of current transmutation doctrines. Further, it is logical that Judy will probably believe, as Levy hypothesizes, that the painting, house, and $100,000—all property she inherited—belong to her because it derived from her inheritance, i.e., separate property.[98] Thus, she will be unaware that placing the painting in the house subjected it to the family-use doctrine. Additionally, Judy's ignorance of Mississippi law will rapidly turn to confusion when she discovers the home she inherited is now subject to both the commingling and family-use doctrines. Finally, she will be unaware that placing the $100,000 into a marital account—albeit for a short period of time—subjected those funds to the commingling doctrine. Unfortunately, Judy will discover Mississippi's transmutation laws much too late when she consults with her attorney after John has filed for divorce. The result of transmutation in Judy's case is staggering. If the chancellor finds that all the inherited property has become marital, the marital estate will be increased by more than $300,000, all of which derived solely from Judy's inheritance.

---

[96] *See, e.g.*, Graham Hughes, *Common Law Systems*, *in* FUNDAMENTALS OF AMERICAN LAW 18-19 (Alan B. Morrison ed., 1996) ("[A legal system] must protect settled expectations which would be dangerously disturbed if the highest courts were free to abandon positions that they had earlier declared with authority."). The author also points out that "[a]ppellate judges in common law systems are very conscious of their responsibility to present lower courts with opinions that offer guidance in future cases, and part of the common law tradition consists of an effort to fashion appellate opinions with this goal in view." *Id.* at 19.

[97] Robert J. Levy, *An Introduction to Divorce-Property Issues*, 23 FAM. L.Q. 147, 152 (1989). Levy continues that certain exceptions to marital property were carved out under the Uniform Marriage and Divorce Act, which included: "(1) assets brought to the marriage by either spouse or assets that [could] be traced directly back to such assets" and "(2) assets acquired by one of the spouses other than by earnings during the marriage." *Id.* at 151-52 (quoting C. FOOTE, K. LEVY, & F. SANDERS, CASES AND MATERIALS ON FAMILY LAW, STATUTORY SUPPLEMENT 651 (3d ed. 1985)). "[This] dichotomy between marital and nonmarital property can be defended by referring to 'essentially common-sense extrapolations of fairness notions and *beliefs about spouses' expectations*.'" *Id.* (emphasis added). Levy also suggests that the dual-classification of property probably aligns better with the parties' expectations of property distribution. *Id.* at 152. *See also* 1 BRETT R. TURNER, EQUITABLE DISTRIBUTION OF PROPERTY § 2:9, at 82-83 (3d ed. 2005) ("There is considerable common sense in the notion that property acquired during the marriage is 'ours,' while property acquired from extramarital sources is 'his' or 'hers.'").

Those that favor the status quo would argue this potential injustice can be cured because the *Ferguson* factors allow the chancellor to adjust the property division settlements, and that the division of marital property does not have to be equal.[99] However, each chancellor has a different view on how significant a separate property contribution will be to the ultimate division of property.[100] One chancellor may find that the injustice is intolerable and adjust the property accordingly, whereas another chancellor may find adjustment is not required. Additionally, the result more than likely will not be remedied at appeal because of the deference given to chancellors for property division issues.[101]

The solution to such a problem is to align Mississippi's laws with those of the parties' expectations. To reach this desired result, Mississippi should abolish the family-use doctrine in its entirety—a result that would instantaneously create clear and consistent precedent for chancellors to follow. This change in jurisprudence would eliminate the arbitrary decisions of lower-level chancellors while additionally enforcing a level of fairness in chancellor's decisions through meeting the parties' expectations. Thus, in John and Judy's case, the home would be classified as separate property, but John would be entitled to his marital share of funds that were contributed to the mortgage through the doctrine of tracing discussed below. Though anecdotal, Judy's scenario demonstrates the family-use doctrine in Mississippi runs contrary to each party's expectations and therefore needs to be abandoned.

---

[98] *See supra* note 97 and accompanying text.

[99] *See, e.g.*, Singley v. Singley, 846 So. 2d 1004, 1012 (Miss. 2002) (providing chancellor does not have to make equal division of marital property).

[100] *See supra* notes 72-75 and accompanying text. Turner also notes that "the identity and personal beliefs of the trial court judge are often a more important factor than the substantive content of the law." 3 TURNER, *supra* note 3, app. A at 255.

[101] *See* Johnson v. Johnson, 650 So. 2d 1281, 1285 (Miss. 1994) ("[An appellate court's] scope of review in domestic relations matters is limited under the . . . rule that [the court] will not disturb a chancellor's findings unless manifestly wrong, clearly erroneous, or if the chancellor applied an erroneous legal standard.").

Next, Mississippi should decisively remove all confusion regarding conflicting commingling doctrines and opt instead for the more ubiquitous method of dual classification used by other states.[102] Interestingly, Mississippi has already adopted dual classification status to various assets, such as pensions[103] and businesses,[104] but has failed to extend this classification to other assets. An extension of dual classification for all assets and the corresponding clearer precedent it would bring would align with each party's expectations and promote fairness in our judiciary system.[105] Thus, in Judy's case, the $100,000 would remain separate property despite the fact it was commingled with marital funds. The length of time it remained in the marital account would be irrelevant and not subject to judicial scrutiny. Further, Judy's house would not be transmuted into marital property due to a small portion of marital money being contributed to an outstanding mortgage balance. Rather, John would be reimbursed for his marital share that was contributed to the mortgage.[106]

---

[102] Turner states that "[a]ll other dual classification states hold that an asset acquired with both marital and separate funds is both marital and separate property." 3 TURNER, *supra* note 3, app. A at 252. Thus, "if 70% of the price to purchase an asset is paid with marital funds, the 70% of the asset is marital property." *Id.* Turner also provides that "[n]o other jurisdiction in the United States presently follows [Mississippi's commingling doctrine]." *Id.* at 254. "In every other state in which the concept of separate property is recognized, property acquired with both marital and separate contributions has simultaneous marital and separate interests." *Id.* For a further discussion of dual classification and tracing doctrines, *see* Deborah H. Bell, *Equitable Distribution: Implementing the Marital Partnership Theory Through the Dual Classification System*, 67 MISS. L.J. 115, 154-55 (1997).

[103] *See, e.g.*, Dye v. Dye, 22 So. 3d 1241, 1245 (Miss. Ct. App. 2009) (providing that the portion of a retirement account that accumulates before marriage is separate property whereas the portion that accumulates during marriage and before divorce is marital).

[104] *See, e.g.*, Pittman v. Pittman, 791 So. 2d 857, 865-66 (Miss. Ct. App. 2001) (husband's business classified as mixed property rather than entirely marital).

[105] *See supra* note 97 and accompanying text.

Mississippi's current family-use and commingling doctrines—along with the puzzling exceptions for each doctrine—should be revised. Currently, Mississippi's family-use doctrine may or may not apply, depending on the chancellor's sole discretion. Such arbitrary and differing decisions leave parties perplexed on whether their property is subject to transmutation. Abolishing the family-use doctrine remedies this bewilderment by conclusively providing that such property is separate regardless of to what degree and the period of time the family utilizes the property. Further, adopting dual classification to all assets combats unfairness and aligns the parties' expectations with the judiciary's outcome.

B. Social Policies

Phasing out the family-use doctrine and implementing dual classification would also significantly enhance various social policy goals of Mississippi.[107] For example, let us return to the Judy example, and now assume that Judy is knowledgeable of Mississippi transmutation doctrines before she inherited the property. Judy's possession of the inherited assets creates two conflicting options, both of them less than ideal. First, Judy can place the expensive painting above the mantle and move her family into the home she inherited from her mother. Regrettably, the consequences of Judy's actions increase the risk that both items will be included in the marital estate, but allow Judy and her family to enjoy the property she inherited. Judy's second option is to conscientiously ensure that both items are never used by her family thereby ensuring neither asset will be included in the marital estate.[108] The latter option protects Judy's property interest in each item but results in an impractical and inefficient use of property—property that should be enjoyed by its owner. If Judy desires to ensure the property remains separate property she will forego option one and proceed under option two. As seen through this example, the "fundamental social policy problem with the family use doctrine is that it changes the behavior of spouses who own separate property."[109] Consequently, current Mississippi law encourages and rewards selfish behavior, and simultaneously provides a strong incentive to individuals such as Judy to hoard or isolate the property from her family.[110]

---

[106] "The strong modern rule is that the marital estate receives not only the marital contributions, but also any growth in the marital contributions after they are made, most often computed through some form of allocation formula." Brett R. Turner, *Unlikely Partners: The Marital Home and the Concept of Separate Property*, 20 J. AM. ACAD. MATRIMONIAL LAW. 69, 90 (2006). In this way, the "marital and separate estates are joint *investors* in the home." *Id.* (emphasis in original). "Each estate is entitled to a fair share of the actual return earned by the investment [and] neither estate [is] compelled to make an interest-fee [sic] loan to the other." *Id.* Mississippi currently recognizes various allocation formulas for pensions and business valuations. *See* BELL, *supra* note 7, § 7.04[5], at 208, § 8.03[3], at 221, § 8.03[3][b], at 222.

[107] The "vague principles" Mississippi currently uses have the benefit of judicial flexibility in determining the outcome of cases. ROBERT E. EMERY, RENEGOTIATING FAMILY RELATIONSHIPS: DIVORCE, CHILD CUSTODY, AND MEDIATION 67 (1994). However, as the author points out, numerous negative consequences arise from this added flexibility, which include: encouragement of litigation because cases are unpredictable;

Abolishing the family-use doctrine would ameliorate the harmful consequences created by this doctrine. Instead of hoarding property, individuals in Mississippi would be encour-

---

animosity amongst the parties is potentially increased because of damaging evidence introduced; judicial bias; and limited appellate review. *Id.*

[108] Judy actually has a third option available—a post-nuptial agreement. A post-nuptial, however, will probably not be signed due to the awkwardness and inconvenience of requesting one from her husband.

[109] *See* Turner, *supra* note 106, at 88.

[110] *See supra* notes 76-80 and accompanying text. "The reward/penalty issue highlights in stark terms why the family use doctrine is so rarely adopted by state courts." Turner, *supra* note 106, at 89. "It is socially desirable that married persons act generously and not selfishly." *Id.* "The family use doctrine inflicts financial harm upon spouses who make a generous sacrifice, while allowing spouses who selfishly refuse to make that sacrifice to retain a benefit." *Id.* "The net effect of the doctrine is not only to encourage selfish refusal to share separate property, but indeed to penalize generous sharing of separate property." *Id.*

aged to use their property for the benefit of their family and themselves. In our example, Judy would not have to evaluate the potential risks of the property transmuting into marital property against the benefit of using that property because the property would be definitively separate. This result would occur regardless of the degree or length of time to which her family used the painting or the home. As an added benefit, Mississippi law would no longer require chancellors to listen to arguments brought forth by both sides in a futile attempt to determine precisely how much time is required to transmute separate property into marital. Abolishment of the family use doctrine would create a clear and consistent precedent chancellors could follow.

C. Difficulty Settling the Case

Finally, and probably most troubling, Mississippi's current transmutation laws encourage litigation.[111] Paradoxically, under Mississippi's current jurisprudence, both sides are armed with conflicting, yet controlling, precedent to use at trial. If the losing party finds the trial unsuccessful, they have ample incentive to attempt to persuade the appellate court that the lower level chancellor erred in his or her interpretation of the law.[112] This incentive to litigate can have a profound effect on the number of cases that eventually proceed to trial and appeals. More importantly, the litigation incentive can have profound effects on the children of the parties.[113] For example, it is now recognized by commentators that the "more fractious the proceeding, the more difficult the aftermath for everyone involved," including children.[114]

---

[111] *See supra* notes 81-82.

[112] *See supra* Part II.

[113] *See* Marsha B. Freeman, *Love Means Always Having to Say You're Sorry: Applying the Realities of Therapeutic Jurisprudence to Family Law*, 17 UCLA WOMEN'S L.J. 215, 217 (2008). Specifically, studies have demonstrated that the societal effect of litigious divorce encourages increased juvenile delinquency, higher teen drop-outs in schools, pregnancy, and effects in marriage rates. *Id.*

[114] *Id.*

Proponents of the status quo would recognize and argue that Mississippi's transmutation laws allow flexibility for chancellors in determining the value of the marital estate.[115] After all, chancellors can use one precedent, such as *Brock* to recognize tracing,[116] or another precedent such as *Boutwell* which broadly provides that commingling converts separate property assets to marital.[117] In effect, chancellors are given wide discretion to decide what they deem to be the fairest division in potentially unlimited and complex circumstances. However, eliminating the family-use doctrine and allowing dual character classification in all instances of commingling would still give chancellors discretion in the *Ferguson* factors;[118] it would only eliminate or narrow what the chancellors can divide through elimination of separate property items as marital assets. Furthermore, any perceived injustice after the property division can be remedied through alimony.[119] Thus, the end result can still be accomplished, but practitioners can better plan their cases with steadfast and concrete laws regarding the property division stage. Alimony may remain a contentious and litigious affair for the parties, but the property division aspect would not. The result, at least in theory, is more settlements and less litigation.[120]

---

[115] *See supra* note 107. *See also* Levy, *supra* note 97, at 149. Judges are required to "choose between clear and predictable rules of law and the discretion which allows them to accomplish individualized justice in particular cases." *Id.*

[116] *See supra* notes 34-39 and accompanying text.

[117] *See supra* note 17 and accompanying text.

[118] Ferguson v. Ferguson, 639 So. 2d 921 (Miss. 1994). Under *Ferguson*, the courts are to look at

> "(1) substantial contribution to property accumulation, including direct or indirect economic contribution, contribution to marital and family stability, and contribution to the education or training of the wage-earning spouse; (2) spousal use or disposition of assets and distribution by agreement; (3) the market and emotional value of assets; (4) the value of each spouse's separate estate; (5) tax consequences and legal consequences to third parties; (6) the extent to which property division can eliminate the need for alimony; (7) the needs of each spouse; and (8) other factors which should be considered in equity."

BELL, *supra* note 7, § 6.01[5][b][iii], at 154.

Thus, even though the marital estate would be smaller, upon making a division of marital property, the chancellor could consider these other factors in making an unequal division of property.

[119] Mississippi recognizes four types of alimony which include: permanent, lump sum, rehabilitative, and reimbursement. BELL, *supra* note 7, § 9.02, at 237. The topic of alimony is beyond the reach of this article, but chancellors could use permanent, rehabilitative, or lump sum alimony as a means towards achieving fairness. For example, if

Thus, elimination of the family-use doctrine would actually encourage settlements and avoid contentious litigation because property once subject to the family-use doctrine would no longer be part of the marital estate. With firm precedent established, the parties would be unable to argue the degree to which certain items of property were used for family purposes, and the separate property would remain separate as a matter of law. Further, elimination of the commingling doctrine in favor of tracing would also reduce the incentive for parties to pro-

---

one party had significant income and a significant separate property estate, and the other spouse had little to no assets, the chancellor could award a higher permanent alimony award or lump sum payment to the destitute spouse. Professor Bell recognizes this potential achievement of fairness in her treatise stating that "lump sum alimony appears to function as an equitable award to a spouse who made substantial long-term contributions to the marriage, but whose contributions cannot be rewarded through property division. *See id.* § 9.02[2][a], at 240. Furthermore, separate property assets can be used in determining the proper amount of alimony received by the recipient. *See, e.g.*, Johnson v. Johnson, 650 So. 2d 1281, 1287 (Miss. 1994); *see also* BELL, *supra* note 7, § 9.04[4][g][ii], at 256 ("Even though a spouse's separate assets cannot be divided in equitable distribution, they may serve as the source for alimony payments.").

[120] "[T]he net effect of [having] unitary property is to shift difficult issues of tracing from the mandatory procedures of the classification stage to the discretionary procedures of the division stage." 1 TURNER, *supra* note 97, § 5:20, at 337. This phenomenon can be seen in various Mississippi cases where the courts recognized "unequal divisions of separate property" when separate property was transmuted into marital through commingling. *See* 3 TURNER, *supra* note 3, at 258 (citing Stewart v. Stewart, 864 So. 2d 934 (Miss. 2003) (husband awarded 80% of marital home that was his separate property before transmutation)). "Practitioners under unitary property quickly learned that some judges would make unequal divisions, others would not make unequal divisions, and many were almost impossible to predict." 1 TURNER, *supra* note 97, § 5:20, at 337. "[This] lack of consistency among cases with similar facts was itself a significant problem, but its more important effect was to make separate property issues very difficult to settle." *Id.* "Because so fewer cases settled, unitary property in actual practice was much *harder* to administer than mixed property." *Id.* "The blunt lesson was that when separate property issues are resolved through the mandatory procedures of the classification stage, the result becomes sufficiently predictable that many contested cases settle before trial." *Id.*

ceed to court. Instead of having multiple cases supporting both sides, there would be established precedent favoring dual classification of assets.

VI. COMPARISON TO OTHER EQUITABLE DISTRIBUTION STATES

Mississippi's adoption of the family-use doctrine derives its genesis from the majority implied gift rule.[121] This majority doctrine—unlike Mississippi's family-use doctrine—requires a specific finding of intent on the part of the donor, either express or implied.[122] Mississippi's family-use doctrine, on the other hand, focuses on one aspect—the use of the property by both spouses or spouse's family, without inquiring into the intent of the donor.[123]

The majority of states require at least some showing of intent by the donor in order to transmute separate property into marital property.[124] For example, in Alaska, "[s]eparate property can become marital property *where that is the intent of the owner and there is an act or acts which demonstrate that intent.*"[125] Thus, "parties may, by their actions during marriage, demonstrate intent to treat specific properties as joint holdings even though they were acquired by one spouse prior to marriage."[126] "*Where such intent is established*, usually through joint management and control of the property, [the] court will consider the property to be a marital asset."[127] *"[T]he act of comingling, in itself, does not automatically establish intent to jointly hold property* . . . ."[128] In contrast, a minority of courts have implied donative intent when the donor places separate property items into the marital household—a doctrine Mississippi seemingly follows.[129]

---

[121] BELL, *supra* note 7, § 6.04[1], at 163, § 6.04[1][b], at 164.

[122] *Id. See also* Bell, *supra* note 102, at 163-64. "Transmutation by implied gift is deemed to occur when an owning spouse, by actions or statements or both, indicates an intention to make a gift of separate property to the marital estate." *Id.* at 163. Courts, in making their determination, have focused on statements made, "the manner in which the property was used during the marriage, and the extent to which the property was commingled with marital funds." *Id.* at 163-64.

[123] BELL, *supra* note 7, § 6.04[1][b], at 163-64. Professor Oldham notes that "[m]ost states thankfully have not adopted this terrible policy, which punishes generous spouses for allowing the other spouse to use separate property." J. Thomas Oldham, *Tracing, Commingling, and Transmutation*, 23 FAM. L.Q. 219, 246 (1989). Oldham continues that "[a]most all married couples permit each other to use the owning spouse's separate property merely as a courtesy." *Id.* at 247. For example, "[o]ne can allow a roommate, guest, friend, or a family member to use property and no change of ownership rights occurs; why should spouses be different?" *Id.* "It would be absurd to require a spouse to place separate property in storage or to bar the other spouse from ever using the property in order to avoid transmutation." *Id.* Oldham concludes, like Turner, that "[t]his would reward selfish and antisocial behavior." *Id.*; *see supra* notes 76-80 and accompanying text.

[124] *See, e.g.*, Sparks v. Sparks, 233 P.3d 1091, 1095 (Alaska 2010) ("The inquiry into whether the property was transmuted required the court to weigh factual evidence of conduct showing an intent to make a donation to the marital unit."); Cuda v. Cuda, 906 S.W.2d 757, 759 (Mo. Ct. App. 1995) ("Non-marital property may lose its character as such if there is evidence of an intent to contribute the property to the community . . . ."); Johnson v. Johnson, 372 S.E.2d 107, 110 (S.C. Ct. App. 1988) ("[T]ransmutation is a matter of intent to be gleaned from the facts of each case."); Coppola v. Coppola, 938

As this comment has recommended, Mississippi should follow other states that require some indication of intent on the part of the donor before conclusively transmuting separate property into marital. As mentioned by Professor Oldham and Brett Turner, the family-use doctrine as it exists today in Mississippi goes contrary to the parties' expectations of property division and against numerous social goals Mississippi should strive to accomplish.[130] These laudable goals can only be accomplished by removing Mississippi from the minority—abolishing the family-use doctrine—and recognizing separate

---

A.2d 786, 791 (Me. 2007) ("[T]ransmutation rule provides that nonmarital property may be transformed into marital property when the spouse . . . exercises an objectively manifested intent to transform the property into marital property."); *see also supra* note 122 and accompanying text.

[125] Chotiner v. Chotiner, 829 P.2d 829, 832 (Alaska 1992) (emphasis added).

[126] Carlson v. Carlson, 722 P.2d 222, 224 (Alaska 1986).

[127] *Id.* (emphasis added).

[128] *Id.* (emphasis added). The opinion, in referring to "comingling" is referring to placing separate property with marital property, not commingling as referred to in this comment in Part II.A. In contrast, the sole act of "comingling" does convert separate property into marital, with some exceptions.

[129] *See, e.g.,* Trattles v. Trattles, 376 N.W.2d 379, 383 (Wis. Ct. App. 1985) (holding household items purchased with separate property funds transmuted into marital property due to use by both spouses); Quinn v. Quinn, 512 A.2d 848, 853-54 (R.I. 1986) (holding inherited furnishings lost their status as separate due to thirty years of use by both spouses).

[130] *See supra* notes 76-80, 123 and accompanying text. *See also* BELL, *supra* note 7, § 6.04[1][b], at 164.

property only transmutes into marital when the intent of the donor is conclusively shown.[131]

Additionally, Mississippi's classification of separate property and marital property as entirely marital places it in a minority. Currently, the "overwhelming nationwide majority rule" has been to classify property with both marital and separate components as mixed property.[132] Other states such as Virginia and Illinois experimented and adopted commingling doctrines as they exist in Mississippi today, but both states eventually abolished the doctrine in favor of dual character classification.[133] The concerns discovered in Virginia and Illinois echo the concerns raised in this comment. Of primary concern was the fact "that too many separate assets [were] mixed together with modest amounts of marital property, so that the practical effect of the doctrine was to destroy too much separate property."[134] As was discovered, "[r]eal people simply [did] not keep marital and separate property sufficiently distinct from one another for unitary property to work as a doctrine."[135] The net effect "was a series of unacceptable problems."[136] "Too much separate property was lost due to minor commingling, too many inconsistent results were reached on the facts of similar cases, and experienced family law practitioners found that it was too difficult to settle cases out of court."[137] To be fair, Illinois ran into additional and unforeseeable problems of their own when they adopted a dual classification of assets regime and thus a change in Mississippi law should be carefully crafted.[138] Accordingly, Mississippi should turn to other states that have found dual classification of assets not only workable but successful.

---

[131] The author advocates that a presumption should be created providing that the property is separate property if the spouse owning the non-marital property can show, through reliable evidence, that the property was obtained before the marriage, by gift or inheritance, or with separate property funds. This presumption can be rebutted through clear and convincing evidence the donor intended to gift the property to the marital estate. Through these standards, the courts will avoid frivolous arguments about donative intent by the non-owning spouse. Additionally, the courts will not be called upon to engage in overly subjective attempts to define intent. Further, the standard is fair because the spouse attempting to prove the property is separate property must demonstrate through evidence that the property was originally separate property. If the spouse is unable to do so, then the standard that "[a]ll assets . . . are presumed to be marital" applies. *See* Wise v. Wise, 37 So. 3d 95, 99 (Miss. Ct. App. 2010).

[132] 1 TURNER, *supra* note 97, § 5:20, at 332.

[133] *Id.* at 333. Illinois courts adopted a unitary theory of property division in 1981, which was subsequently abolished by state legislature in 1983. *Id.* Virginia "adopted unitary property in 1987" and the Virginia legislature reversed the doctrine in 1990. *Id.* at 333-34.

[134] 3 TURNER, *supra* note 3, app. A at 201.

[135] *Id.*

[136] *Id.* at 369.

The Virginia legislature—unhappy with the consequences of unitary property—proposed a solution by codifying a definition of "hybrid" or dual classification property.[139] Under the statute, chancellors faced with commingled property are required to treat the property as "part marital property and part separate property" unless the degree of commingling is to such an extent that the originating funds are no longer identifiable.[140] The statute further provides:

> When marital property and separate property are commingled by contributing one category of property to another, resulting in the loss of identity of the contributed property, the classification of the contributed property shall be transmuted to the category of property receiving the contribution. However, to the extent the contributed property is retraceable by a

---

[137] *Id.*

[138] *Id.* at 370.

[139] VA. CODE ANN. § 20-107.3 (2008). More specifically, the statute requires the court to "determine the legal title as between the parties, and the ownership and value of all property . . . and shall consider which of such property is separate property, which is marital property, and which is part separate and part marital property in accordance with subdivision A 3." *Id.* § 20-107.3 (A). "The dual classification of property or 'hybrid' property is a concept that recognizes the ability of the court to take an asset and generally classify it into two component parts, which parts recognize the premarital and postmarital contributions to the asset by either party." PETER N. SWISHER, LAWRENCE D. DIEHL & JAMES R. COTTRELL, VA. PRAC. FAMILY LAW § 11.6 (11th ed. 2010). Most importantly, "[i]t recognizes the accumulation of value of an asset as an ongoing process, and when property is acquired with an expenditure of both marital and separate property, it is classified as part marital and part separate." *Id.* This method is "the majority theory in equitable distribution states." *Id.*

[140] VA. CODE ANN. § 20-107.3(A)(3) (2008). "The requirement to make a hybrid classification of property by the trial court is mandatory, [and therefore] removes the discretionary inconsistency and lack of guidance that had previously resulted from . . . prior case law." SWISHER ET AL., *supra* note 139, § 11.06.

> preponderance of the evidence and was not a gift, such contributed property shall retain its original classification.[141]

Virginia's statute has been lauded as a "success" and "remarkable for its directness and simplicity."[142] A review of the statute reveals reasons for its success. First, the statute instructs chancellors to classify commingled property as mixed property, and thus chancellors are prohibited from and will be reversed if they fail to follow the guidelines.[143] Additionally, courts are instructed on proper procedure to determine the classification of property that has been commingled to such an extent that it has lost an identifiable status. In Virginia's case, the property that is contributed transforms into the underlying receiving property's classification. To avoid this classification, a party must demonstrate to the court by a preponderance of evidence that the property is traceable. For example, return to the hypothetical scenario of Judy.[144] Under Virginia's statute, the residence would be classified as separate property because it was inherited. When John contributes either separate property funds of his own or marital funds to pay off the remaining mortgage, and the couple proceeds to divorce, the home will be classified as a mixed asset. On the other hand, if the money John contributed to the equity of the home was so commingled with the underlying asset, the cash John contributed would become the separate property of Judy. For John to avoid this unfortunate consequence, he would have to demonstrate by a preponderance of the evidence that he can retrace the money contributed to the mortgage to his separate account or marital account. John can do this through receipts, testimony, or other various ways. This may seem confusing to the reader, but it is easily applied in Virginia courts and has the added benefit of laying down a firm foundation for chancellors to work with. However, the beauty of Virginia's statute is not in the substantive parts created for division of property, but rather in the consistent rules provided that judges must follow.

---

[141] VA. CODE ANN. § 20-107.3(A)(3) (2008).

[142] *See* 3 TURNER, *supra* note 3, app. A at 369, 371.

[143] *See* Duva v. Duva, 685 S.E.2d 842, 846-47 (Va. Ct. App. 2009) (reversing trial court's determination that property was marital when evidence was introduced husband owned residence prior to marriage).

[144] *See supra* Part I.

Many other equitable distribution states have adopted what is commonly referred to as the source of funds doctrine as a remedy to the harsh unitary theory of property practiced by Mississippi courts.[145] Under this doctrine:

> A spouse contributing non-marital property is entitled to an interest in the property in the ratio of the non-marital investment to the total non-marital and marital investment in the property. The remaining property is characterized as marital property and its value is subject to equitable distribution. Thus, the spouse who contributed non-marital funds, and the marital unit that contributed marital funds each receive a proportionate and fair return on their investment.[146]

This doctrine has various advantages that would combat the disadvantages realized by Mississippi's current commingling doctrine. First, the source of funds doctrine encourages, rather than penalizes, parties to contribute assets to the marital estate. Through this doctrine, fears are relieved that courts may one day classify mixed property as entirely marital. The doctrine also aligns better with policies of fairness. Both estates are compensated for their investment into the asset. If the asset increases in value, both estates realize a return on their investment. Finally, it promotes rather than discourages parties to settle. Both parties are well aware that the court will classify the asset as mixed and divide the asset accordingly. Parties are not left to wonder whether the asset has been commingled to such a degree that the chancellor may or may not classify the asset as entirely marital or worse, that the chancellor believes some exception to the commingling doctrine applies.

---

[145] *See* Glaspy v. Glaspy, 545 S.E.2d 782, 787 (N.C. Ct. App. 2001) (applying source of funds rule to real property); Goodman v. Goodman, 760 N.E.2d 72, 79 (Ohio Ct. App. 2001) ("[D]octrine of transmutation has been replaced by the concept of traceability, i.e., the 'source of funds' rule."); Hoffmann v. Hoffmann, 676 S.W.2d 817, 825 (Mo. 1984) (adopting source of funds rule).

[146] Thomas v. Thomas, 377 S.E.2d 666, 669 (Ga. 1989).

Both the Virginia method and the source of funds method use tracing for determining the proper classification of commingled property. It is noteworthy that tracing has not gone without judicial criticism. In a concurring opinion, Judge Vance of the Kentucky Court of Appeals called tracing "a complicated legal and accounting process . . . inimical to the spirit of marriage."[147] He continued that "[m]arriage connotes sharing, with the concentration upon what is 'ours' rather than what is 'his' and what is 'hers.'"[148] Judge Vance then concludes that "[i]t does not bode well for the institution of marriage if each partner must keep in the back of his mind the possible advantage to be obtained by keeping up with and being able to trace every penny brought into the marriage."[149]

Tracing doctrines can become admittedly difficult to administer, but Mississippi's current all-or-nothing approach to commingling is arguably more complicated. This explains Mississippi's appellate court decisions applying tracing in some situations but readily holding in others that tracing does not apply.[150] In an effort to promote fairness, the appellate courts have fashioned a rather rigid case-by-case approach where tracing may or may not apply. Further, Judge Vance's characterization of tracing is misplaced. Tracing potentially strengthens marriages because spouses are more readily willing to commit separate funds into the marital estate. In response to Judge Vance's final conclusion, the author would submit that those worried about "tracing every penny" are precisely the individuals who under Mississippi law worry about transferring separate property to the marital estate for fear the property will be converted into all marital property. Thus, like Judge Vance hypothesizes, those individuals in Mississippi that do not worry about transmutation issues will not suddenly become encouraged through implementation of a tracing doctrine to reevaluate their financial decisions.

---

[147] Turley v. Turley, 562 S.W.2d 665, 669 (Ky. Ct. App. 1978) (Vance, J., concurring).

[148] *Id.*

[149] *Id.*

[150] *See supra* notes 34-41 and accompanying text.

As illustrated throughout this comment, division of marital property continues to be a challenging issue for many courts and properly explains why there are various methods used to determine the proper classification of property. However, the purpose of this comment is not to advocate one position over another, but rather to illustrate that better possibilities exist for this state beyond the status quo. Mississippi's current methods have been tried in two other states and have conclusively failed in those states. Whatever method Mississippi eventually attempts to implement, the main focus of such a change needs to concentrate around rules that are both clear and consistent.

VIII. CONCLUSION

Mississippi's current transmutation laws need major revisions, and these revisions should come by way of clear and concrete precedent. Establishing a dual-property classification system; allowing tracing in all situations but those where assets are no longer readily identifiable; and eliminating the family-use doctrine will result in a fairer judicial system. Instead of maintaining several various precedents which allow judges to arbitrarily pick and choose, a more reliable system for all parties will result in laws that are consistent and predictable. This predictability and consistency aids attorneys and their clients in deciding the strength or weakness of their cases and facilitates settlement as both sides know their respective legal positions. The steadfast rules of consistent precedent also lead to less arbitrary decisions by Mississippi courts and therefore greater assurance from the public in the judiciary. Finally, the parties' expectations will align with those of the relevant laws, which results in a fairer division to separate property owners.

Matthew L. Roberts[*]

---

[*] The author would like to thank Professor Deborah H. Bell, Professor of Law, University of Mississippi School of Law, for her guidance in preparing this comment. This comment is dedicated to my grandmother, Nell Parker, who recently passed away.

COMMENT

NURTURING THE UNITED STATES COMMERCIAL SPACE INDUSTRY IN AN INTERNATIONAL WORLD: CONFLICTING STATE, FEDERAL, AND INTERNATIONAL LAW

Because outer space is accessible from everywhere on Earth, the law of outer space must address activities extending into space regardless of where those activities originate. States have created both national and international law to address the legal concerns involved in space activity. While international law is crucial to coordinating this effort among states, states, as sovereigns, ultimately have control over what occurs in the areas within their jurisdiction. Thus, while the goals and obligations of international space law are important, a foundation of national law is also crucial in order to ensure that it is effective in practice. National law can reflect local interests, thus giving citizens more of a stake in space activity that operates smoothly under both national and international law. However, when local interests and international interests appear to conflict, states have to contend with the ambiguity of international space law as they attempt to comply with their obligations to both their citizens and the international community.

In the United States, this national space law was first created, like international space law, as a reaction to and facilitator of the political climate. Since then, the United States' primary goals have shifted to focus on development of an economically viable commercial space sector. Both state and national governments have worked within their respective legal spheres to facilitate local interests while complying with the law of the larger legal groups of which they are a part. While the United States' interests have evolved, international law has remained the same, and the tension between their respective goals has been heightened by ambiguity concerning the ways in which the national and international law that represent those goals are related. This paper argues that, from a United States national law point of view, this ambiguity is not as problematic as it seems, and might actually create legal ambiguities that are helpful to both the national and state governments who are attempting to invest in and profit from the commercial space industry. First, this article presents an overview of the current state and federal legal regimes designed to nurture the commercial space industry. By examining the relationship between national law and its purposes, international law and its purposes, and local law and its purposes, it goes on to argue that this complex web of law creates an equally complex web of potential federal preemption of national law. This complexity ensures that the fundamental concerns and policy goals of state, national, and international law and actors are all represented and gives decision-makers the tools to make the appropriate decisions when called upon to do so.

I. STATE RESPONSIBILITY UNDER ARTICLE VI OF THE OUTER SPACE TREATY

The Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies, commonly known as the Outer Space Treaty (OST),[1] starts from the principle that space exploration and use should benefit all, regardless of "degree of . . . economic or scientific development"; that space should be used for peaceful purposes; that cooperation between states should be encouraged in order to promote international understanding; and that the general "purposes and principles of the Charter of the United Nations" should be incorporated into the otherwise specifically space-based purposes and principles of the treaty.[2] The treaty proceeds in Articles I – XIII to outline these principles more concretely.[3]

---

[1] Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, *opened for signature* Jan. 27, 1967, 18 U.S.T. 2410, 610 U.N.T.S. 205 [hereinafter Outer Space Treaty].

In the commercial space context, in which private actors perform tasks that the OST seems to assume will be performed by states, questions arise as to how much responsibility States Parties have for the actions of their citizens and those who launch within their territory, in addition to what actions they must take domestically in order to satisfy their obligations under international law.[4] Article VI of the OST provides that:

> States Parties to the Treaty shall bear international responsibility for national activities in outer space, including the Moon and other celestial bodies, whether such activities are carried on by governmental agencies or by non-governmental entities, and for assuring that national activities are carried out in conformity with the provisions set forth in the [Outer Space] Treaty. The activities of non-governmental entities in outer space, including the Moon and other celestial bodies, shall require authorization and continuing supervision by the appropriate State Party to the Treaty.[5]

Further, as mentioned above, the OST provides that a State Party to the treaty is liable for damage caused if it "launches or procures the launching of an object into outer space" and "from whose territory or facility an object is launched" for damage caused by such an object.[6]

---

[2] *Id.*

[3] Designating outer space as "the province of all mankind," Article I requires non-discriminatory access to the moon and other celestial bodies. *Id.* at art. I. Article II forbids state appropriation of these bodies. *Id.* at art. II. The treaty also stresses the importance of international cooperation and the freedom of scientific investigation, *id.* at arts. IX, X, XI.; forbids national appropriation of "[o]uter space, including the moon and other celestial bodies," *id.* at art. II.; forbids placing nuclear weapons in orbit around Earth, *id.* at art. IV.; requires states and their astronauts to provide assistance to the astronauts of other states, *id.* at art. V.; imposes liability for damage caused by the space objects of one State Party to the Treaty to another State Party to the Treaty or "its natural or juridical persons," i*d.* at art. VII.; and requires registration of those objects and obligates states to return the registered objects of other states found within their territories, extending ownership of objects into outer space. *Id.* at art. VIII.

[4] Ronald L. Spencer, Jr., *State Supervision of Space Activity*, 63 A.F. L. REV. 75, 79-84 (2009); *see also* Frans G von der Dunk, *Sovereignty Versus Space—Public Law and Private Launch in the Asian Context*, 5 SING. J. INT'L & COMP. L. 22, 30-34 (2001) (discussing sources of jurisdiction which enable nations to exercise control over their citizens and thus ensure that those citizens comply with international law).

[5] Outer Space Treaty, *supra* note 1, at art. VI.

In terms of responsibility, because States Parties are the only legal actors, they are the only entities with the power, on an international level, to fulfill the obligations created by the OST and accompanying treaties. Thus, they are the entities that will be held responsible for a failure to do so. Note, however, that the States Parties are only responsible for the "national activities" carried on by non-governmental entities. Because "national activities" are undefined, States Parties are left to define it as they see fit and to act accordingly.[7] This ambiguity reflects the fact that the key players in international law are sovereigns, battling and compromising on national interests in a way that ultimately allows each to maintain its sovereignty by allowing all to define a crucial phrase in the treaty. The United States has created its own national regulatory regime that reflects both how it interprets these obligations and other national policy goals.

II. FEDERAL COMMERCIAL SPACE LEGISLATION

A. Commercial Space Launch Act of 1984

When Congress enacted the Commercial Space Launch Act of 1984 (CSLA),[8] it sought to encourage the private sector to invest in space technology in order to "complement the launching and associated services . . . available from the United States government,"[9] "promote economic growth and entrepreneurial activity,"[10] and "enable the United States to retain its competitive position internationally."[11] The CSLA was meant to further these efforts by instituting "stable, minimal, and appropriate regulatory guidelines"[12] and to create a more streamlined and efficient licensing application system.[13] Regulation was to be implemented only "to the extent necessary" to ensure all private activity complied with United States international obligations and to provide for the "public health and safety, safety of property, and national security interests and foreign policy interests of the United States."[14]

---

[6] *Id.* at art. VII.

[7] P.J. Blount, *3rd Eilene Galloway Symposium on Critical Issues in Space Law – Panel: Article VI: The Legal Landscape*, Dec. 11, 2008, http://rescommunis.wordpress.com/2008/12/11/3rd-eilene-galloway-symposium-on-critical-issues-in-space-law-panel-article-vi-the-legal-landscape/ (summarizing Joanne Gabrynowicz, Director, National Center for Remote Sensing, Air, and Space Law, Presentation at 3rd Eilene Galloway Symposium on Critical Issues in Space Law: Overview of States Responsibility in a Global Commons (Dec. 11, 2008)); *see also* von der Dunk, *supra* note 4, at 28-30.

[8] Commercial Space Launch Act, Pub. L. No. 98-575, 98 Stat. 3055 (1984) (codified as amended at 49 U.S.C. §§ 70101-70121 (2006)) [hereinafter Commercial Space Launch Act of 1984].

[9] *Id.* § 2(4).

The power to issue licenses and oversee commercial launch activities—and thus to ensure that these guidelines were met—was concentrated in the Department of Transportation (DOT).[15] The Secretary of the Department was given exclusive federal licensing authority[16] and tasked with regulating in a manner consistent with the themes running throughout the Act: balancing the interests of the private sector and the government, facilitating their shared interest in efficiency and a streamlined licensing process, and encouraging extensive consultation between the Secretary of Transportation and other executive agencies.[17] In order to carry out these purposes, the Secretary was given broad discretion to grant licenses, waive license requirements, and suspend, revoke, modify, or terminate licenses based on a determination that to do so was in the "public interest."[18] The Secretary was required, however, to coordinate with other agencies to ensure consistency of licensing requirements, make certain that all interests were represented, and guarantee that health, safety, and national policy and security concerns were properly considered.[19] Additionally, the Secretary's discretion was subject to administrative and judicial review[20] and tempered by notification[21] and consultation requirements,[22] as well as by decision deadlines.[23]

---

[10] *Id.* § 3(1).

[11] *Id.* § 2(5).

[12] *Id.* § 2(6).

[13] *Id.* § 3(2); 130 CONG. REC. 14,900 (1984) (statement of Mr. Fuqua).

[14] Commercial Space Launch Act of 1984, *supra* note 8, at § 2(7).

[15] 130 CONG. REC. 14,900 (1984) (statement of Mr. Fuqua).

[16] Commercial Space Launch Act of 1984, *supra* note 8, at §§ 6(c)(1), 7.

[17] *See id.* §§ 3(2), (3), 5(2), 16, 20; *see also* 130 CONG. REC. 14,900 (statement of Mr. Fuqua).

[18] *See* Commercial Space Launch Act of 1984, *supra* note 8, at §§ 8(c), 10(a).

The CSLA forbade any person to "launch a launch vehicle or operate a launch site within the United States, unless authorized by a license issued or transferred under [the CSLA]."[24] Further, it forbade any United States citizen from launching a vehicle or operating a launch site without "a license issued or transferred under [the CSLA]" in any location.[25] For the purposes of the Act, "United States citizen" was defined as "any individual who is a citizen of the United States" or "any corporation, partnership, joint venture, association, or other entity organized or existing under the laws of the United States or any State . . . or . . . the laws of a foreign nation, if the controlling interest . . . in such an entity is held by an individual or entity" that qualifies as a citizen under one of the other definitions of "citizen."[26]

In addition to licensing regulations, the Act directed the Secretary "to facilitate and encourage the acquisition (by lease, sale, . . . or otherwise) by the private sector of launch property . . . or . . . services . . . of the United States" that were not being used for public purposes.[27] It further set out costs for lease and sale of Government facilities. The greater of fair market value or value of the benefit constituted the sale price; and direct costs, including diminishing value of property and costs of trained personnel paid as a result of the use, dictated the price for other types of transfer, including lease.[28] The Secretary was further directed to prescribe an amount of liability insurance to require of each licensee in order to meet U.S. international obligations.[29] The Act proceeded to specify a civil penalty of not more than $100,000 for each violation, with a new violation occurring daily.[30] Finally, the Act clarified its relationship to state and other law.[31] The Act explicitly stated that

---

[19] *Id.* §§ 5(a)(2), 8(a)(2). Federal agencies were also required to assist the Secretary to the fullest extent possible. *Id.* § 5(b). Thus, the Secretary was as much a centralized point through which regulation could be coordinated as he or she was a regulatory authority.

[20] *Id.* § 12.

[21] *See, e.g.*, *id.* § 10(d).

[22] *See id.* § 20.

[23] *See id.* § 9(b).

[24] *Id.* § 6(a)(1).

[25] *Id.* § 6(a)(2), (3).

[26] *Id.* § 4(11).

[27] *Id.* § 15(a).

> [n]othing in this Act shall preclude a State or a political subdivision of a State from adopting or putting into effect any law, rule, regulation, standard, or order which is consistent with this Act and is in addition to or more stringent than any requirement of or regulation issued under this Act[,]

and authorized the Secretary to collaborate with states to make the process of approval of space launch activities as efficient and simple as possible.[32] Further, the Act made clear that a launch vehicle does not become an export for the purposes of export law simply by being launched and that none of the provisions of the Act applied to U.S. Government activities.[33] Finally, the CSLA required that the Secretary carry out all obligations in a manner that complied with any treaty, convention, or agreement in force between the United States and a foreign nation.[34]

B. 1988 Commercial Space Launch Act Amendments

In 1988, Congress amended the Commercial Space Launch Act to include a risk-sharing liability scheme. This scheme covered liability for claims against licensees by other parties involved in launch activities, as well as by third parties.[35] Further, it clarified the meaning of "direct costs"[36] and expanded the services for which the government could charge to include "activities involved in the production of a launch vehicle or its payload for launch if such activities were agreed to by the owners or manufacturers of such launch vehicle or payload."[37] These Amendments responded to a National Security decision directive, which was, in turn, a reaction to the *Challenger* tragedy, by ending NASA's launch role for commercial and foreign satellites.[38] With the sole U.S. provider of launch services excluded from the market, American satellite operators were forced to turn to the "fledgling U.S. commercial launch industry or launch[] on foreign vehicles."[39]

---

[28] *Id.* § 15(b)(1).

[29] *Id.* § 16.

[30] *Id.* § 19.

[31] *Id.* § 21.

[32] *Id.* § 21(a).

[33] *Id.* § 21(b), (c).

[34] *Id.* § 21(d).

[35] Commercial Space Launch Act Amendments of 1988, Pub. L. No. 100-657, § 5, 102 Stat. 3900 (codified as amended at 49 U.S.C. §§ 70101-70121 (2006)). "Third party"

Thus, the primary purpose of the 1988 Amendments was to make the United States more competitive in what had become an international market for launch services.[40] In particular, it was designed to encourage U.S. customers, who had contracted with NASA and were left without means through which to launch their payloads, to launch within the United States.[41] By offering a liability regime similar to the one that had applied to NASA launches, and to that offered by the United States' largest competitor, the French Arianespace, Congress hoped to encourage the private sector to invest in domestic launch property and services.[42] To this end, the Amendments not only provided one-time incentives for the satellite providers whose satellites NASA would not be able to launch as

---

was defined as "any person or entity other than (A) the United States, its agencies, or its contractors or subcontractors involved in launch services; (B) the licensee or transferee; (C) the licensee's or transferee's contractors, subcontractors, or customers involved in launch services; or (D) any such customer's contractors or subcontractors involved in launch services." *Id.* § 3.

[36] *Id.* § 4(b) ("[T]he term 'direct costs' means the actual costs that can be unambiguously associated with a commercial launch effort, and would not be borne by the United States Government in the absence of a commercial launch effort.").

[37] *Id.* § 4(d).

[38] 134 CONG. REC. 12,112 (1988) (statement of Mr. Roe).

[39] *Id.*

[40] *Id.*

[41] 134 CONG. REC. 12,113 (1988) (statement of Mr. Roe).

[42] *Id.*

planned,[43] but also recommended that the United States initiate an international dialogue aimed at developing "reasonable and fair international competition in . . . space activities."[44] Finally, signaling the level of importance it attached to the commercial launch industry through its willingness to invest in the industry's future, and drawing a parallel to the government-industry relationship in the field of aeronautics, Congress "direct[ed] NASA to support research into launch systems component technologies" and "to consult with [the] industry in order to facilitate the transfer of technology to the commercial sector."[45]

The risk-sharing regime was intended not only to encourage private sector investment but also to ensure that the United States complied with its international obligations under the Outer Space Treaty and the Convention on International Liability for Damage Caused by Space Objects.[46] Congress was well aware that the United States, as a "launching state,"[47] would be responsible for any damage the private sector might cause to the space objects or spacecraft of other parties to the treaty. Because it was in the national interest to allow the private sector to enter the space launch business, Congress was willing to place some of the risk on the U.S. Government.[48]

The regime required that every licensee "obtain liability insurance or . . . demonstrate financial responsibility, in an amount sufficient to compensate the maximum probable loss . .

---

[43] *See* Commercial Space Launch Act Amendments of 1988, Pub. L. No. 100-657, § 6, 102 Stat. 3900 (codified as amended at 49 U.S.C. §§ 70101-70121 (2006)).

[44] *Id.* § 9.

[45] 134 CONG. REC. 12,113 (1988) (statement of Mr. Roe).

[46] Convention on International Liability for Damage Caused by Space Objects, *opened for signature* Mar., 29 1972, 24 U.S.T. 2389, 961 U.N.T.S. 13810 [hereinafter Liability Convention].

[47] The OST provided that a State Party to the treaty that "launch[ed] or procure[ed] the launching of an object into outer space" and "from whose territory or facility an object [was] launched" would be internationally liable to a state party to the treaty to whom damage is caused by its space object. Outer Space Treaty, *supra* note 1, at art. VII. The Liability Convention defines such a state as the "launching state," and states that a "launch" includes an intended launch. Liability Convention, *supra* note 46, at art. I(b), (c).

[48] *See* 134 Cong. Rec. 12,112 (1988) (Statement of Mr. Roe) (discussing the proposed allocation of risk between the launch provider and the U.S. Government).

. from claims by a third party for death, bodily injury, or loss of or damage to property resulting from activities carried out under the license";[49] and from claims by the United States "for loss of or damage to property of the United States."[50] However, Congress instituted a liability cap of $500,000,000 for third party claims and $100,000,000 for U.S. claims, or "the maximum liability insurance available on the world market," whichever was the lesser amount.[51] The Amendment further required reciprocal waivers of claims between licensees and their contractors, subcontractors, and customers, and the contractors and subcontractors of those customers.[52] Each of these parties thus agreed to absorb the costs of any damages that might occur to its property or injuries its employees might sustain.[53] Further, the United States waived any perspective claims against the licensee in excess of the maximum probable loss.[54] The United States also agreed to pay successful claims "for death, bodily injury, or loss of or damage to property" by third parties against a licensee, if those claims were not covered by the maximum probable loss insurance and if the United States was given the opportunity to control or assist in the defense of the claim.[55] All parties would be liable for any damages that occurred due to willful misconduct.[56]

The Amendments further emphasized the importance of cooperation between the Secretary of Transportation and other government agencies, particularly the Department of Defense and NASA, reflecting a realization that licensing alone was not sufficient to create a stable environment necessary to encourage the investment Congress felt the United States desperately needed.[57]

---

[49] Commercial Space Launch Act Amendments of 1988 § 5(a).

[50] *Id.*

[51] *Id.*

[52] *Id.*

[53] *Id.*

[54] *Id.*

[55] *Id.* However, the United States would only cover liability in excess of maximum probable loss up to 1.5 billion dollars. *Id.*

[56] *Id.*

[57] *Id.*

C. 1991 National Aeronautics and Space Administration Authorization Act

The National Aeronautics and Space Administration Authorization Act for the 1991 Fiscal Year amended the CSLA to reflect a new interest in creating a "strong space transportation infrastructure with significant private sector involvement."[58] Further, the Act encouraged state governments to facilitate private sector involvement to establish the launch sites, as well as the complementary and support facilities, of which that infrastructure would be composed.[59] As the private sector became more active in the launch industry, NASA became more involved with the private sector, and Congress invested more in the space industry, establishing more extensive oversight capabilities in order "to ensure that the taxpayers' money [was] spent as effectively as possible to achieve the space goals Congress [had] adopted."[60] Recognizing the vital contribution of the "high-risk long-term" aeronautics research program at NASA to the preeminence of U.S. aerospace programs, Congress expressed the intent to promote similar preeminence in the space industry by continuing to fund such research.[61] Additionally, Congress extended the purposes and findings of the CSLA to stress the importance of developing commercial space transportation infrastructure and to enlist the states' help in establishing that infrastructure.[62]

D. Commercial Space Act of 1998

The Commercial Space Act of 1998 expanded the licensing regime to cover reentry sites and gave the Secretary the power to issue safety procedures governing launch and reentry.[63]

---

[58] National Aeronautics and Space Administration Authorization Act, Fiscal Year 1991, Pub. L. No. 101-611, § 117(c)(3), 104 Stat. 3188 (1990) [hereinafter 1991 NASA Auth. Act].

[59] *Id.*

[60] 136 CONG. REC. 34,155 (1990) (statement of Mr. Sensenbrenner).

[61] 136 CONG. REC. 34,156 (1990) (statement of Mr. Lewis).

[62] 1991 NASA Auth. Act, *supra* note 58, at § 117(d)(3).

[63] Commercial Space Act of 1998, Pub. L. No. 105-303, § 102, 112 Stat. 2843 (codified as amended at 49 U.S.C. §§ 70101-70121 (2006)).

Congress instituted this change to encourage investment in reusable launch vehicles, which it predicted would be the future of the space industry.[64] While stressing the fact that the federal government should continue "to undertake those activities that only it can and/or should carry out," Congress expressed a strong commitment to the efficiency, and freedom from government competition, which would facilitate the development of the commercial space sector. [65]

E. 2004 Commercial Space Launch Act Amendments

The most recent amendments to the Commercial Space Launch Act occurred in 2004 and instituted safety and licensing requirements "[t]o promote the development of the emerging commercial human space flight industry."[66] Additionally, the amendments expanded the Secretary's licensing power and gave the Secretary power to issue experimental permits.[67] The amendments were needed for several reasons—the fledgling commercial human space flight industry was unable to plan for its future due to the absence of a licensing regime for the reusable launch vehicles that it was developing; the existing licensing regime covered only expendable launch vehicles (ELVs), which did not return to Earth intact; and safety and liability standards were needed to protect the safety of the humans who participated in the development of human space flight technology.[68]

Central to the 2004 Amendments was an attempt to balance the need to deter risk through regulation with the desire to leave the private sector the freedom to innovate and create new technology.[69] The Amendments created two new classes of people: "crew" defined as employees of a licensee or licensee's subcontractor who are directly involved in the launching or reentry of a vehicle that carries human beings, and "space flight participant[s]" defined as individuals who are not crew but are carried in space flight vehicles.[70]

---

[64] 143 CONG. REC. 24,346 (1997) (statement of Mr. Rohrabacher).

[65] 143 CONG. REC. 24,348 (1997) (statement of Mr. Brown).

[66] Commercial Space Launch Amendments Act of 2004, Pub. L. No. 108-492, 118 Stat. 3974 (codified at 49 U.S.C. §§ 70101-70121 (2006)).

[67] *Id.* § 2(b)(6).

[68] H.R. REP. NO. 108-429, pt. 2, at 2-3 (2004).

[69] Commercial Space Launch Amendments Act of 2004 § 2(a)(5).

The Amendments required a licensee to inform both crew and space flight participants in writing, before making a contract or launching human beings into space, "that the United States Government [had] not certified the launch vehicle as safe for carrying crew or space flight participants."[71] The licensee was directed to inform a space flight participant of the "risks of launch and reentry, including the safety record of the launch or reentry vehicle type"; the Secretary was required to inform a space flight participant of information concerning risk or probable loss during the flight.[72] Crew members were required to receive training. Finally, both groups were subject to medical or other standards promulgated by the Secretary.[73]

The Secretary was further allowed to issue regulations "governing the design or operation[s] of a launch vehicle to protect the health and safety of crew and space flight participants."[74] However, those regulations would only apply to launches in vehicles that had ended in a serious or fatal injury or "contributed to an unplanned event or series of events during a licensed or permitted commercial human space flight that posed a high risk of causing a serious or fatal injury."[75]

Much of the Congressional debate focused on concerns about the fact that no safety regulations were required until a fatal accident occurred. In the words of Congressman Oberstar, "Is the gentleman going to include on the space flight ticket the disclaimer there has been no safety provided until after you are dead?"[76] Congressman Rohrabacher countered this with the contention that people, especially in a developing industry like human space flight, have the right to waive safety requirements, which they might not waive if the industry were more developed.[77] This debate reflected the broader attempt to "strike a delicate balance between the need to give a new industry a chance to develop brand-new technology and the desire to provide enough regulation to protect the industry's customers."[78] In the end, that balance was manifested in an informed consent requirement, which was bolstered by the Secretary's ability to regulate pre-flight health and training requirements; this requirement was supplemented by a hands-off approach that allowed for industry innovation up until the point that death or a high level of risk would lead to the overwhelming preeminence of safety concerns.[79]

---

[70] *Id.* § 2(b)(2), (9).

[71] *Id.* § 2(c)(13).

[72] *Id.*

[73] *Id.*

[74] *Id.* § 2(c)(14).

[75] *Id.*

[76] 150 CONG. REC. H10,048 (daily ed. Nov. 19, 2004) (statement of Mr. Oberstar).

To offset the limitation on innovation these safety requirements might impose, the Amendments introduced an experimental permit option.[80] The Secretary was given the power to issue permits for reusable suborbital rockets that would be launched or reentered solely for research and development, demonstrate compliance with licensing requirements, and train crew before obtaining a license "using the design of the rocket for which the permit would be issued."[81] Further, they stressed the Secretary's authorization to waive requirements of law when issuing permits, subject to considerations of "public health and safety, safety of property, and national security and foreign policy interests of the United States."[82]

The 2004 Amendments reflected a continued dedication to nurturing a growing private space sector through increased congressional oversight by requiring the Department of Transportation to streamline what the industry considered "overly burdensome and unnecessary regulations" that did not "allow for the rapid development of new and innovative launch vehicle designs" and "prolonged and painstaking licensing process[es] that place[d] excessive financial burdens on applicants."[83] They also constituted an acknowledgement that if the space sector was to operate in a manner consistent with public interest, increased innovation in the human space flight industry required a reciprocal increase in safety regulation.

---

[77] 150 CONG. REC. H10,050 (daily ed. Nov. 19, 2004) (statement of Mr. Rohrabacher).

[78] 150 CONG. REC. H10,049 (daily ed. Nov. 19, 2004) (statement of Mr. Boehlert).

[79] *See id.*

[80] Commercial Space Launch Amendments Act of 2004 § 2(b)(6).

[81] *Id.* § (2)(c)(16).

[82] *Id.*; *see also* H.R. REP. NO. 108-429, pt. 7, at 8 (2004).

[83] H.R. REP. NO. 108-429, pt. 7, at 8. (2004).

The CSLA and Amendments created a regulatory environment that limited liability, provided stability and predictability, and struck a balance between safety and innovation. The legislation thus limited the risk of investment in the commercial space industry in a way that made it feasible for additional private entities to become involved, obtain insurance, and make a profit. In so doing, this regulatory environment made it potentially profitable to states to encourage investors to locate in their jurisdictions.

III. STATE INCENTIVES FOR THE COMMERCIAL SPACE INDUSTRY

A. General Incentives

States offer a variety of tax, zoning, and other incentives to attract businesses that will create jobs and other economic benefits.[84] States are especially interested in attracting businesses with high-paying jobs like the commercial space transportation sector. Many states offer the commercial space sector not only traditional incentives but also the typically higher incentives offered to high-technology businesses. These include tax rebates or exemptions; tax incentives aimed at particular industries, such as high technology tax credits, rewards for providing jobs that are high-paying relative to the local median income, research and development tax credits, and exemption or advantage zones designed to encourage an increase in the standard of living, employment, or population in a specific area of the state.[85] Additionally, some companies are attracted by university or industry development zones that encourage a particular industry and university to collaborate, often by sharing facilities, and are frequently accompanied by local incentives.[86]

---

[84] FED. AVIATION ADMIN., STATE SUPPORT FOR COMMERCIAL SPACE ACTIVITIES 1 (2008).

[85] *Id.* at 5.

[86] *Id.* at 7.

A few states offer incentives specific to the space transportation industry. These usually take the form of infrastructure-providing entities like space authorities, which are often staffed by people educated in the field who serve as contact points and advocates within the state, and space ports.[87] Additionally, they may have bonding authority that allows them to finance infrastructure that would be extremely expensive for a launch provider to provide on its own.[88] Moreover, some states have recently begun enacting informed consent and tax incentives aimed at the space transportation industry.[89]

B. Spaceports

The Webster-Merriam Dictionary Online defines "spaceport" as "an installation for testing and launching spacecraft,"[90] reflecting the fact that the term "spaceport," as the extraterrestrial counterpart to the airport, is commonly associated specifically with transportation. The Federal Aviation Administration (FAA) defines "spaceport" as a launch and reentry site.[91] For the purposes of this paper, the FAA definition will be controlling. A "launch site," as defined by the CSLA, is "the location on Earth from which a launch takes place . . . and necessary facilities at that location;"[92] a "reentry site" is "the location on Earth to which a reentry vehicle is intended to return."[93] Most state law incorporates these definitions, either implicitly or explicitly, into its spaceport legislation.

Spaceports offer a range of services, which can vary not only with the facilities they offer, but also as a result of environmental constraints.[94] Spaceports may include launch pads, runways, access to restricted airspace, payload processing and integration, telemetry, tracking, and fuels, as well as infrastructure that allows them to provide launch preparation services and to serve space flight participants and crew.[95] A spaceport may simply rent or lease the facilities, provide complete launch services, or offer any group of services in between. The type of services a spaceport provides is governed by the type of customers it is designed to serve. Spaceports generally either specialize in supporting reusable launch vehicles (RLVs), space flight participants (SFPs), and horizontal take-off and landing vehicles (HTOLs); or, alternately, in supporting expendable launch vehicles (ELVs).[96]

---

[87] *Id.* at 7-8.

[88] *Id.* at 7.

[89] *Id.* at 3-4.

[90] Webster-Merriam Dictionary Online, http://www.merriam-webster.com/dictionary /spaceport (last visited Mar. 29, 2010).

[91] FED. AVIATION ADMIN., 2010 U.S. COMMERCIAL SPACE TRANSPORTATION DEVELOPMENTS AND CONCEPTS: VEHICLES, TECHNOLOGIES, AND SPACEPORTS 61 (2010) [hereinafter 2010 FAA Report].

[92] 49 U.S.C. § 70102(7) (2006).

[93] *Id.* § 70102(15).

The FAA contrasts commercial spaceports with federally-owned and -operated "ranges."[97] When ranges are used exclusively for the purposes of the federal government, they are exempt from CSLA licensing requirements and regulations. However, while ranges were originally created to serve exclusively federal needs, the CSLA specifically provides for lease, as well as potential sale, of these facilities and the services they provide, and they are now frequently used by commercial customers. [98] Further, state-run spaceports are often co-located with federal ranges, taking advantage of their well-established infrastructure.[99]

C. Overview of State Spaceport Law

Although state spaceport legislation varies somewhat, it is united in its attempt to balance the interests of the community with those of the business it is seeking to attract. That being said, the level of specificity, the complexity, and the relationships to the government and other organizations specific to each state's spaceport regime reflects the unique history and goals of each.

---

[94] Michael C. Mineiro, *Law and Regulation Governing U.S. Commercial Spaceports: Licensing, Liability, and Legal Challenges*, 73 J. AIR L. & COM. 759, 761 (2008).

[95] *Id.*; 2010 FAA Report, *supra* note 91, at 61.

[96] Mineiro, *supra* note 94, at 761.

[97] 2010 FAA Report, *supra* note 91, at 61-62.

[98] Mineiro, *supra* note 94, at 762.

[99] 2010 FAA Report, *supra* note 91, at 69.

i. California

Given California's size and history, it is perhaps unsurprising that its space incentive regime is complex, involving cooperation between multiple agencies. The two main actors supporting the commercial space sector are the Business Transportation and Housing Agency (BTH) and the California Space Authority.[100] The state's Space Enterprise Development Act directs the BTH to "foster activities that increase the competitiveness of space enterprise in California, including, but not limited to, the commercial use of space, space vehicle launches, space launch infrastructure, manufacturing, applied research, technology development, economic diversification, and business development."[101] In order to achieve these goals, the BTH is authorized to contract with other entities, both private and public, and required to select a California nonprofit corporation to help it achieve its goals.[102] The nonprofit is to serve as the California Spaceport Authority; to provide information and recommendations to the secretary; and to "[d]evelop and implement a state strategy for applying and commercializing technology" in order to improve the state economy and foster innovation within the industry.[103] A wide variety of entities, from cities to special districts to private entities, can apply to the authority for designation as a spaceport, provided the applying entity receives a license from the FAA and provides the Authority with notice of license approval.[104] However, spaceport designation will be revoked for failure to comply with FAA safety requirements.[105]

California includes spaceports within its legislation facilitating the formation of airport districts, which require cost-sharing among the cities and counties that benefit from the spaceport or airport establishment.[106] Districts may sue and be sued, acquire real or personal property necessary or convenient to carry out their purposes, including acquisition through condemnation, and issue and sell bonds to fund a spaceport or an airport.[107] Finally, a Spaceport Office has been established within the Department of Transportation "to seek and obtain federal funding for the commercialization of private space activities in California."[108] Clearly, California considers the commercial space industry worthy of investment.

---

[100] FED. AVIATION ADMIN., *supra* note 84, at 12.

[101] CAL. GOV'T CODE § 13999.2(a) (West 2005).

[102] § 13999.2(b), (c).

[103] § 13999.2(d)(4).

[104] § 13999.3(a), (b).

[105] § 13999.3(d).

ii. New Mexico

New Mexico's location makes it an obvious candidate for space-related activity. Not only does it offer a range of traditional tax advantages, but its history of aerospace technology development and the presence of White Sands Missile Range make it particularly unique.[109] Virgin Galactic is already slated to be the anchor tenant of the state's Spaceport America.[110] The New Mexico Regional Spaceport District Act creates a spaceport authority and classifies it as a state agency.[111] Combinations of governmental units form contracts to establish regional spaceport districts, the Authority issues certificates authorizing the combination to exercise the statutorily granted powers of a district, and the district becomes a separate political subdivision of the state with issuance of the certificate.[112] A district may, among other things, establish a regional spaceport within its boundaries, enter into contracts and agreements affecting its affairs, and sue and be sued.[113] Further, a district may contract with the spaceport authority for the authority to issue bonds to fund spaceport-related endeavors.[114] These bonds mature within a maximum of twenty to thirty years.[115] A municipality or county that wants to become a member of a regional spaceport district must impose a municipal or county regional spaceport gross receipts tax for the privilege of doing business in the state, but only after a majority of the members of the governing body in a municipality approve the tax.[116] Receipts from performing integral spaceport-related activities, such as launching, operating and recovering space vehicles, may be deductible, as well as receipts from operating a spaceport and preparing a payload.[117] A minimum of seventy-five percent of the proceeds then go to spaceport development.[118] Overall, New Mexico's complex financing scheme facilitates not only infrastructure development, but also local involvement in the spaceport development process.

---

[106] CAL. PUB. UTIL. CODE § 22002 (West 2005).

[107] § 22553.

[108] CAL. GOV'T CODE § 14007.2(a) (West 2005).

[109] FED. AVIATION ADMIN., *supra* note 84, at 20.

[110] *See* Todd G. Dickson, *Spaceport Begins Work on Virgin Galactic Hangar*, LAS CRUCES BULL., Mar. 26, 2010, http://www.mveda.com/blog/2010/03/spaceport-begins-work-on-virgin-galactic-hangar/ (describing the progress made on the hangar being built for Virgin Galactic's use and providing general information about the spaceport, job creation, and New Mexico's new informed consent statute).

[111] N.M. STAT. § 58-31-4(A) (2006).

[112] *Id.* § 5-16-4(A).

iii. Florida

Florida legislation establishes Space Florida, an independent special district of the state.[119] The stated goals of the legislation, "to foster the growth and development of a sustainable and world-leading aerospace industry in [the] state . . . by facilitating business financing, spaceport operations, research and development, workforce development, and innovative education programs,"[120] are clearly manifested in the powers of Space Florida. Space Florida may, among other things, sue and be sued; "acquire, enjoy, use and dispose of" intellectual property; lend money; invest and reinvest its funds; and "[h]ave and exercise all powers necessary or convenient to effect any or all of the purposes for which it is organized."[121] In order to carry out its purposes, Space Florida is given the ability to establish charges to pay for the development of spaceport facilities. Further, it is given the power to pledge contracts as security for bonds, which also help to finance spaceports, as well as to further the other goals of Space Florida.[122] The primary goals of the Florida space program are reflected in the definitions contained in the statute, which are less concerned with defining all of the services and events that could take place at a spaceport, and more with clarifying the means through which those services and events will be offered and take place. Thus, the definitions focus not on clarifying the meaning of "launch" or "reentry," but on defining "business client," "person," and "project."[123]

---

[113] *Id.* § 5-16-6.

[114] *Id.* § 58-31-6(A); § 5-16-7.

[115] *Id.* § 58-31-6(C)(3).

[116] *Id.* § 7-19D-15.

[117] *Id.* § 7-9-54.2.

[118] *Id.* § 7-19D-15(B), (D).

[119] FLA. STAT. ANN. § 331.302 (West 2003).

[120] § 331.302(1).

iv. Texas

Texas legislation creates spaceport development corporations and gives counties and combinations of municipalities and counties the power to authorize the creation of such corporations.[124] A spaceport development corporation may exercise eminent domain power but must first receive approval from the governing body of the county or city in which the relevant property is located.[125] Such a corporation also has the capacity to contract and is required to contract out of any liability due to environmental damage.[126] Property, income, and operations of a spaceport development corporation are exempt from state and local taxes, while tangible personal property in the spaceport is exempt from ad valorem taxation.[127] However, ad valorem taxes are paid by a corporation to the political subdivisions of the state in the amount that would be paid on the land owned by the corporation if it was owned by a private entity.[128] A spaceport is designated as a special type of "project" and is thus subject to the law governing projects in addition to that which specifically regulates spaceports.[129] Overall, Texas law is characterized by its ambiguity, and thus by the discretion it offers to local governments to develop spaceports in a manner so as to both facilitate development and take into account local concerns and interests.

---

[121] § 331.305(4), (7).

[122] FLA. STAT. ANN. § 331.305(20) (West Supp. 2009).

[123] *See* FLA. STAT. ANN. § 331.303 (West 2003).

[124] TEX. LOC. GOV'T CODE ANN. §§ 507.001, 507.003 (Vernon Supp. 2009).

[125] *Id.* § 507.103.

[126] *Id.* § 501.104.

[127] *Id.* § 507.201.

v. Oklahoma

Oklahoma law establishes the Oklahoma Space Industry Development Authority, a body corporate and politic, with all of the essential rights and privileges of the government.[130] The Authority, a government agency, is charged with overseeing and facilitating the development of spaceports and spaceport facilities,[131] which are required to be located within specific parts of the state.[132] It has the power to obtain up to twenty-five acres of land at a time through eminent domain and to collect revenues such as tolls and fees to help fund facilities and services; enter into contracts; and use those contracts as security for bonds.[133] The Authority thus has broad power within the designated spaceport territory to act on issues relevant to the operation, maintenance, and safety of a spaceport.[134] The statute also establishes a Spaceport Territory Advisory Council consisting of local representatives which may assist and advise the Authority, at the Authority's discretion, on matters related to the establishment, operation, and maintenance of facilities, as well as make recommendations on issues of local concern, such as land use and development, zoning, and safety regulations.[135] Very similar to Florida's regime, the primary difference is Oklahoma's inclusion, albeit minimal, of local government through the Advisory Council.

---

[128] *Id.* § 507.202.

[129] *Id.* § 507.002

[130] OKLA. STAT. tit. 74, § 5203(A) (2009).

[131] *Id.* § 5203(B).

[132] *Id.* §§ 5202(17), 5213.

[133] *Id.* § 5204.

[134] *Id.*

[135] *Id.* § 5212.

vi. Alaska

The Alaska Aerospace Development Corporation is a public state corporation,[136] the purpose of which is to promote economic development, assist the state in becoming a lead player in space exploration and development, and to facilitate space-related educational and economic growth.[137] The Corporation is particularly directed to promote use of the Poker Flat Research Range as a launch and scientific research site and has a close relationship to the University of Alaska.[138] The Corporation has the power to sue and be sued, make contracts, and perform other corporate functions in order to fulfill its purpose.[139] It has the power to issue tax-exempt bonds, subject to a million dollar cap.[140] The statute indemnifies the state, the University of Alaska, and board members from liability deriving from a lack of sufficient securities for bonds.[141] The Corporation, as a public agency, is also indemnified from liability for negligence to a construction-related professional service.[142] The statute's straight-forward definitions of the essential components of a spaceport reflect the initial simplicity of Alaska's facilities and focus on research and technology development.[143]

vii. Virginia

Virginia has enacted very detailed legislation describing the powers of its Commercial Space Flight Authority to create projects for the purposes of "disseminat[ing] knowledge pertaining to scientific and technological research and development among public and private entities" and "promot[ing] industrial and economic development."[144] In particular, the Authority is directed to promote and facilitate the national preeminence of the Mid-Atlantic Regional Spaceport.[145] The Authority has the right to sue and be sued, acquire projects and property, issue bonds, and adopt bylaws, among other powers.[146] A failure to define "launch," "spaceport," or any other term typically found in state statutes of this sort, combined with the Authority's authorization to use any legal means necessary to promote its stated goals, reflects the state's broad commitment to facilitating research and economic development in whatever form they may present themselves.

---

[136] Alaska Stat. § 14.40.821 (2008).
[137] *Id.* § 14.40.861.
[138] *Id.* §§ 14.40.861(3)-(4), 14.40.821.
[139] *Id.* § 14.40.866.
[140] *Id.* § 14.40.891.
[141] *Id.* § 14.40.899.
[142] *Id.* § 36.90.300.
[143] *Id.* § 14.40.990.

viii. Hawaii

Less developed than that of other states, Hawaii's legislation indicates an intention to explore the potential for space activity by establishing an Office of Aerospace Development in the Department of Business, Economic Development, and Tourism.[147] The director of this Office is required to "direct the planning, evaluation, and coordination of space-related activities and identify and promote opportunities for expanding and diversifying aerospace-related industries in the State, which may include [establishing] a Pacific international center for space exploration systems to support space exploration and settlement."[148] Additionally the director is to help the University of Hawaii and other local entities form partnerships to facilitate the aerospace industry, as well as to promote education, increase public awareness of aerospace programs, and gather information about the state of the industry in general.[149] An aerospace advisory committee assists the director and is composed of representatives of business, of each of the main islands, and of the public universities.[150] Finally, Hawaii's legislation specifically forbids a vehicle launched from the state from "carry[ing] weapons of destruction, nuclear waste materials, or us[ing] radioactive materials as a power source."[151]

---

[144] VA. CODE ANN. §§ 2.2-2201, -2202 (2007).

[145] *Id.* § 2.2-2699.2(6).

[146] *Id.* §§ 2.2-2204 to -2217.

[147] *See* HAW. REV. STAT. § 201-73 (2001).

[148] *Id.* § 201-73(1).

[149] *Id.* § 201-73(3)-(5).

[150] *Id.* § 201-72.5.

ix. Wisconsin

The Wisconsin Aerospace Authority was created to, among other things, promote, advertise, and develop Wisconsin's aerospace industry, facilitate related educational opportunities, "attract . . . space-related businesses," coordinate access to aerospace services, and promote public use of its spaceport facilities.[152] Wisconsin has instituted a distinctive funding program for "spaceport improvement project[s]," which include "any project to acquire, construct, develop, plan, design, establish, create, improve, enlarge, reconstruct, or equip any spaceport or spaceport facility."[153] The costs of these projects, up to fifty percent of excess costs if it is a federally funded project, or up to eighty percent of total costs if not federal funded, are borne by the Authority and the state.[154] The actual percentage paid depends on the importance of the specific project to the state spaceport program.[155] Municipalities, counties, and other spaceport owners can also apply for an advance land acquisition loan in order to fund the land on which to build a spaceport.[156]

D. Informed Consent Statutes

Virginia, Florida and, most recently, New Mexico have adopted liability regimes as further incentives to commercial space transportation entities, for whom exposure to liability in the fledgling field is a significant concern.[157] Although the

---

[151] *Id.* § 201-75.

[152] WIS. STAT. ANN. § 114.63 (West Supp. 2008).

[153] *Id.* § 114.77.

[154] *Id.* § 114.77(2)-(3). State payment of costs is subject to a ten million dollar cap. *Id.*

[155] *Id.* § 114.77(4).

[156] *Id.* § 114.375.

[157] *See* VA. CODE ANN. §§ 8.01-227.8 to -227.10 (2007); FLA. STAT. ANN. § 331.501 (West Supp. 2009); S.B. 9, 49th Leg., Reg. Sess. (N.M. 2010).

CSLA requires that licensees provide safety information to, and obtain informed consent from, space flight participants, it is ambiguous about exactly what information the space flight participant must be given in order to be sufficiently informed and the extent to which space flight participants waive their claims. State clarification of what commercial space flight entities must do to minimize liability and of the scope of the liability that does exist, better allows those entities to minimize and predict their potential liability.[158] By allowing insurance providers to do the same, such liability regimes may lower the maximum probable loss and thus the cost of insurance.

The informed consent legislation of all three states is very similar. That of New Mexico and Florida, with one difference, is virtually identical. Virginia's is slightly broader, offering more entities protection from more types of liability. All three regimes begin by equating the terms "participant" and "space flight activities" with the terms "space flight participant" and "launch services" or "reentry services," respectively, as those terms are defined by the CSLA.[159] Virginia diverges somewhat by defining "space flight entity." All three states define space flight entity to include public or private entities licensed by the United States Federal Aviation Administration.[160] Virginia also protects public and private entities that indirectly hold such licenses "through a corporate subsidiary or a parent."[161] Virginia differs by including under this definition "any manufacturer or supplier of components, services, or vehicles that have been reviewed by the United States Federal Aviation Administration as part of issuing . . . a license, permit, or authorization."[162] Each statute then goes on to outline the circumstances under which a "space flight entity" will be immune from exposure to liability.[163]

---

[158] FED. AVIATION ADMIN., *supra* note 84, at 1-2.

[159] VA. CODE ANN. § 8.01-227.8 (2007); FLA. STAT. ANN. § 331.501(1)(a), (b) (West Supp. 2009); S.B. 9, § 2(A), (B), 49th Leg., Reg. Sess. (N.M. 2010).

[160] VA. CODE ANN. § 8.01-227.8 (2007); FLA. STAT. ANN. § 331.501(1)(c) (West Supp. 2009); S.B. 9, § 2(C), 49th Leg., Reg. Sess. (N.M. 2010).

[161] VA. CODE ANN. § 8.01-227.8 (2007).

[162] *Id.*

In Florida and New Mexico, space flight entities are not subject to liability "for injury to or death of a participant resulting from the inherent risks of space flight activities," if the statutory warning is given.[164] Thus, a participant or his or her representative cannot seek damages under such circumstances.[165] Virginia's statute explicitly indemnifies space flight entities from liability from claims resulting from "bodily injury, including death, emotional injury, or property damage sustained by a participant . . . if such injury or damage results from *the* risks of the space flight activity."[166] Virginia also provides an illustrative list of possible representatives of the participant who are barred from bringing claims (heirs, administrators, etc.), including a general prohibition forbidding any person from bringing a claim on the participant's behalf and includes a reference to federal law, requiring that a space flight participant be informed of the dangers of space flight activities as required by the CSLA.[167]

All three states deny a space flight entity exemption if the entity's act or omission proximately causes the injury or damages and "constitutes gross negligence evidencing willful or wanton disregard for the safety of the participant;" and if the space flight entity intentionally harms a participant.[168] New Mexico and Florida also deny exemption if the space flight entity "[h]as actual knowledge or reasonably should have known of a dangerous condition on the land or in the facilities or equip-ment used in the space flight activities and the danger proximately causes injury, damage, or death to the participant."[169]

---

[163] VA. CODE ANN. § 8.01-227.9 (2007); FLA. STAT. ANN. § 331.501(2) (West Supp. 2009); S.B. 9, § 3, 49th Leg., Reg. Sess. (N.M. 2010).

[164] FLA. STAT. ANN. § 331.501(2)(a) (West Supp. 2009); S.B. 9, § 3(A), 49th Leg., Reg. Sess. (N.M. 2010).

[165] FLA. STAT. ANN. § 331.501(2)(a) (West Supp. 2009); S.B. 9, § 3(A), 49th Leg., Reg. Sess. (N.M. 2010).

[166] VA. CODE ANN. § 8.01-227.10(B) (2007) (emphasis added).

[167] *Id.*

[168] VA. CODE ANN. § 8.01-227.9(B) (2007); FLA. STAT. ANN. § 331.501(2)(b)(1), (3) (West Supp. 2009); S.B. 9, § 3(B)(1), (3), 49th Leg., Reg. Sess. (N.M. 2010). However, the Virginia exception to indemnification does not include damages caused by negligence, but only to injuries sustained as a result thereof. VA. CODE ANN. § 8.01-227.9(B)(1) (2007).

Finally, each statute contains an informed consent provision containing a warning that must be signed by a participant for a space flight entity to receive indemnity with respect to the injuries of that participant.[170] The warnings of New Mexico and Florida include a simple statement of space flight entity exemption in the case of injury or death that results from the inherent risks of space flight activity.[171] However, their examples of what might constitute an injury differ somewhat. Notably, the New Mexico statute includes emotional injury, while the Florida statute includes injury to land and animals.[172] The Virginia warning similarly restates its indemnity regime, including as part of the warning provisions a requirement that a participant be informed of the risks of space flight as required by federal law, and certify that he or she has been given the opportunity to speak with an attorney before signing the warning and giving his or her consent to the state indemnification regime.[173]

This brief overview of incentives makes it clear that states feel they have an interest in enticing the commercial space industry and are willing to invest in the infrastructure required to attract that industry. However, because this law exists within a larger context of national and international law, its efficacy is dependant not only upon its local effectiveness, but also on the extent to which it fits within larger legal regimes.

IV. PREEMPTION

In the United States, the manner in which state, national, and international law interact is governed by the Constitution.

---

[169] FLA. STAT. ANN. § 331.501(2)(b)(2) (West Supp. 2000); S.B. 9, § 3(B)(2), 49th Leg., Reg. Sess. (N.M. 2010).

[170] VA. CODE ANN. § 8.01-227.10(A) (2007); FLA. STAT. ANN. § 331.501(3)(a) (West Supp. 2009); S.B. 9, § 3(A), 49th Leg., Reg. Sess. (N.M. 2010).

[171] FLA. STAT. ANN. § 331.501(3)(b) (West Supp. 2009); S.B. 9, § 4(A), 49th Leg., Reg. Sess. (N.M. 2010).

[172] FLA. STAT. ANN. § 331.501(3)(b) (West Supp. 2009); S.B. 9, § 4(A), 49th Leg., Reg. Sess. (N.M. 2010).

[173] VA. CODE ANN. § 8.01-227.10(B) (2007).

While from an external point of view, the United States as a nation may be obligated to comply with international law, from an internal point of view, "[i]nternational law is part of our law"[174] and forms a subset of national law. As a part of national law, international law must contend with the implications of the dual sovereignty created by the Constitution. Thus, while seen as controlling on an international level, on the national level, international law may be subject to state law, and the way in which it is implemented may depend upon a complex interaction between multiple sovereigns, structural considerations, like separation of powers, and debates concerning the purpose of the relevant law. The Constitution's Supremacy Clause governs these interactions. In the context of commercial space transportation, which is simultaneously governed by state, national, and international law, knowing which the Supremacy Clause favors in case of a conflict is vital to protecting and encouraging large investments.

A. The Supremacy Clause

The Supremacy Clause states

> [t]his Constitution, and the Laws of the United States which shall be made in Pursuance thereof; and all Treaties made, or which shall be made, under the Authority of the United States, shall be the supreme Law of the Land; and the Judges in every State shall be bound thereby, any Thing in the Constitution or Laws of any State to the Contrary notwithstanding.[175]

The consequent power of the federal government to be "supreme within its sphere of action"[176] was further defined and solidified in *McCulloch v. Maryland* in Chief Justice John Marshall's opinion invalidating the state of Maryland's tax on the Second Bank of the United States.[177] Reasoning that the power to destroy is an essential part of the power to create, the

---

[174] The Paquete Habana, 175 U.S. 677, 700 (1900).

[175] U.S. CONST. art. VI, cl. 2.

[176] McCulloch v. Maryland, 17 U.S. (4 Wheat.) 316, 405 (1819).

[177] *Id.* at 436.

Court made it clear that any state law that restricted the federal government's ability to make the "supreme law of the land" violated the Supremacy clause and impinged upon the rights of other states that the Clause was created to preserve.[178]

In *Gibbons v. Ogden*,[179] the Court further expounded upon the scope of federal power by invalidating a New York law that prohibited trading vessels licensed by the United States from entering waters within the jurisdiction of the state.[180] Thus, the Court established that state law must yield to federal law in the case of a conflict between the two, regardless of whether or not the state exercised its legitimate power in making that law. States were not only forbidden to intrude upon the power specifically granted to the national government by the Constitution but it also lay beyond state power to undermine the law created pursuant to that power.[181]

Since then, the Court has had to grapple with the conflict between the Supremacy Clause and the dual sovereignty created by the Constitution. These early cases laid the groundwork for the many subsequent cases in which courts struggled to find the point at which, in the absence of an explicit statement by Congress, the conflict between federal and state law rose to a level requiring a limitation on the power of the states. The United States Supreme Court has developed three general types of preemption, each of which provides a solution tailored to address a particular type of conflict between state and federal power.

B. The Commerce Clause

The Commerce Clause provides one example of a Constitutional provision that can give the federal government power to legislate in a particular legal area. Case law concerning preemption under the Commerce Clause is thus primarily focused on the determination of whether the Commerce Clause gives Congress the power to legislate on a specific subject.

---

[178] *Id.* at 406.

[179] 22 U.S. (9 Wheat.) 1 (1824).

[180] *Id.* at 86.

[181] *Id.* at 81-82.

> The Commerce Clause gives Congress power to regulate three types of activity: (1) "the use of channels of interstate commerce"; (2) "the instrumentalities of interstate commerce, or persons or things in interstate commerce, even though the threat may come only from intrastate activities"; and (3) "those activities having a substantial relation to interstate commerce, . . . i.e., those activities that substantially affect interstate commerce."[182]

After finding that a challenged federal statute constitutes a valid exercise of Congressional power, the Court will do a normal preemption analysis.

C. Preemption Analyses

Ultimately, "there is not—and from the very nature of the problem there cannot be—any rigid formula or rule which can be used as a universal pattern to determine the meaning and purpose of every act of Congress," and the area of law that act circumscribes.[183] However, regardless of the category of preemption into which a particular statute ultimately falls, the Court generally does a similar analysis. First, the Court determines whether a state statute falls within the federal government's "sphere of influence," within which state law is subordinate to federal law.[184] Then, the Court decides whether that invasion is sufficiently opposed to the objectives underlying the federal law to necessitate preemption.[185] "[T]he purpose of Congress is the ultimate touchstone in every preemption case."[186] Generally, unless the conflict between federal and state law is clear or explicit, the Court will "look to the federal statute itself, read in the light of its Constitutional setting and its legislative history."[187] Further, it begins with a presumption in favor of local law in cases dealing with areas traditionally regulated by states.[188] Finally, the Court has developed three general preemption categories—express, conflict, and field—to guide preemption analysis.[189]

---

[182] United States v. Tykarsky, 446 F.3d 458, 470 (3d Cir. 2006) (quoting United States v. Lopez, 514 U.S. 549, 558-59 (1995)).

[183] Hines v. Davidowitz, 312 U.S. 52, 67 (1941).

[184] McColloch v. Maryland, 17 U.S. (4 Wheat.) 316, 421 (1819) (the powers of the federal government are limited to those which are within the scope of the Constitution).

[185] Hines v. Davidowitz, 312 U.S. 52, 67 (1941) (stating that preemption is appropriate if the court finds that, "under the circumstances . . . [state] law stands as an obstacle to the accomplishment and execution of the full purposes and objectives of Congress").

D. Foreign Policy and Preemption

When a state law presents an obstacle to compliance with a treaty or federal law particularly relevant to international affairs, the Court is especially likely to preempt, particularly when the Executive Branch has indicated its belief that preemption is appropriate.[190] International law is a particularly federal realm, concerning other nations and relevant to the nation as a whole; because foreign policy affects all states, there is a heightened presumption that the federal government, rather than a single state unilaterally, should have the power to determine what that policy will be and how to act on it.[191] In the case of a conflict between a state law and a treaty to which the United States is a party, the treaty is treated essentially like a federal statute, and treaty interpretation may play a significant role in determining whether and to what extent preemption is appropriate.[192]

Perhaps the most famous U.S. case involving conflict between a United States treaty obligation and state law is *Missouri v. Holland.*[193] In that case, the state of Missouri challenged the Migratory Bird Act on Tenth Amendment grounds.[194] The Migratory Bird Act gave effect to an agreement between the United States and Great Britain providing for closed seasons and other protections for specific species of endangered migratory birds.[195] Missouri objected to these regulations, claiming that they invaded the Tenth Amendment powers reserved to the states and that, if Congress lacked the power to regulate in a particular area in the absence of a treaty, then a treaty should not give Congress that power.[196] The Court refused to limit the treaty power of the national government and invalidated the Missouri law because the state had no property claim to the birds and because the relevant species of birds travelled, the states were incompetent to achieve the purpose of the treaty on their own.[197] In the words of Justice Holmes, "[i]f the treaty is valid there can be no dispute about the validity of the statute under Article 1, Section 8, as a necessary and proper means to execute the powers of the Government."[198] Essentially, the power to make treaties gives the federal government power to implement treaties.

---

[186] Wyeth v. Levine, 129 S. Ct. 1187, 1194 (2009).

[187] *Hines*, 312 U.S. at 79 (Stone, J., dissenting).

[188] Rice v. Santa Fe Elevator Corp., 331 U.S. 218, 230 (1947).

[189] Hillsborough County v. Automated Med. Lab., 471 U.S. 707, 712-13 (1985).

[190] *See* Crosby v. Nat'l Foreign Trade Council, 530 U.S. 363, 385-86 (2000).

[191] *See id.* at 377.

[192] *See* El Al Isr. Airlines v. Tsui Yuan Tseng, 525 U.S. 155, 162-68 (1999) (analyzing the Warsaw Convention to determine whether preemption was appropriate by analyzing the plain text, drafting history, and intent of the parties).

[193] 252 U.S. 416 (1920).

When the Court was confronted with the effects of Massachusetts foreign policy on U.S. national foreign policy in *Crosby v. National Foreign Trade Council,* it considered whether those effects frustrated a federal statutory objective.[199] Because the Massachusetts law undermined the ability of the President to exercise authority over economic sanctions against Burma, the Court found that "[the state law] undermine[d] the intended purpose and natural effect of . . . the federal Act."[200] Further, the state sanctions conflicted with the federal sanctions in application and duration.[201] Differentiating means from ends, and stating that agreement concerning the former did not justify an endless range of the latter, the Court invalidated the Massachusetts law.[202]

---

[194] *Id.* at 430-31.

[195] *Id.* at 431.

[196] *Id.* at 432.

[197] *Id.* at 435.

[198] *Id.* at 432. There has been and still is much debate concerning Justice Holmes' assertion, which some scholars believe opens the door to endless expansion of federal legislative power, allowing Congress to regulate any area of law that can be tied to compliance with treaty obligations. *See, e.g.*, Nicholas Quinn Rosenkranz, *Executing the Treaty Power*, 118 HARV. L. REV. 1867 (2005); David M. Golove, *Treaty-Making and the Nation: The Historical Foundations of the Nationalist Conception of the Treaty Power*, 98 MICH. L. REV. 1075 (2000); LOUIS HENKIN, FOREIGN AFFAIRS AND THE UNITED STATES CONSTITUTION 185-98 (2d ed. 1996) (discussing limitations on treaty power).

[199] Crosby v. Nat'l Foreign Trade Council, 530 U.S. 363 (2000).

V. NATIONAL PREEMPTION OF STATE LAW PURSUANT TO THE COMMERCE CLAUSE AND/OR OUTER SPACE TREATY

To analyze the extent to which federal law might require preemption of state law, one must first determine that the federal law is valid: i.e., that it falls within the federal government's "sphere of action" and that it became law through the proper means.[203] The OST and other space treaties to which the United States is a party clearly fall within the federal "sphere of action" as the Supremacy Clause explicitly includes "all Treaties made . . . under the Authority of the United States."[204] Because the United States is a State Party to these treaties, the President has, "with the Advice and Consent of the Senate,"[205] exercised his Article II treaty-making power, making the OST part of the "Law of the Land."[206] However, there might be some question as to whether the space treaties are self-executing, and thus roughly "equivalent to an act of the legislature";[207] or if, in contrast, specific legislative acts are required in order to satisfy the treaty requirements.[208] This issue goes to the heart of the issue concerning the extent of state responsibility under Article VI of the OST, as well as to the question of whether adoption of the OST requires federal preemption of state law.

---

[200] *Id.* at 373.

[201] *Id.* at 376-77.

[202] *Id.* at 378-80.

[203] *See supra* notes 176-81 and accompanying text.

[204] U.S. CONST. art. VI, cl. 2.

[205] U.S. CONST. art. II, § 2, cl. 2.

[206] *Id.* However, "treaties are subject to the constitutional limitations that apply to all exercises of federal power, principally the prohibitions of the Bill of Rights . . . ." HENKIN, *supra* note 198, at 185.

[207] HENKIN, *supra* note 198, at 199.

[208] *Id.* at 199-200.

While the OST is most likely self-executing as applied to federal government activities, it may be non-self-executing in regard to the private space-faring entities within the United States. Whether Congress has an obligation to regulate those entities under the OST turns largely on how the Article VI phrase "national activities" is defined. If "national activities" means only the activities governed by the OST that are performed for or with the involvement of the national government, Congress may not be required to regulate at all; the legislature could simply leave execution of the treaty to the executive branch (which would then be the sole space-faring entity for which the U.S. would be responsible). On the other hand, if responsibility for non-governmental "national activities" includes all activities related to the OST that are performed by non-governmental entities within the nation, or non-governmental national citizens outside of the nation, then Congress has a clear obligation to regulate. Regardless of how this ambiguous phrase is ultimately defined on an international level, there is little doubt that Congress has created legislation in a manner that assumes some type of responsibility for the actions of its citizens and those using launch facilities within the United States (those subject to U.S. jurisdiction). Thus, for the purposes of this paper, it will be assumed that, at least to the extent that they involve actions in which private entities can and do engage, the OST and its progeny are non-self-executing "national activities;" that Congress has executed these treaties, at least to some extent, by enacting the CSLA; and that the OST thus gives Congress a power to regulate which may or may not preempt the power of states to regulate in the same field.

Putting aside the OST for a moment, one can also consider the validity of Congressional law-making in the area of space transportation, even in the absence of a treaty. This power arguably falls within the Commerce Clause power to "regulate Commerce with foreign Nations, and among the several states."[209] Arguably, launched objects intrude into federal airspace, entering the "channels of interstate commerce" and thus becoming "instrumentalities of interstate commerce."[210] Congress's power to enact federal legislation governing air commerce, as well as the legislation defining the national airspace in which that commerce occurs, is derived from the Commerce Clause.[211] Further, if space objects actually go into outer space, they travel into an international commons outside of the territorial jurisdiction of the United States.[212] Thus, in the future, spaceports may be the sites of "activities that substantially affect interstate commerce."[213] Launch sites have the potential to become transportation hubs similar to airports, serving spacecraft on which people travel quickly from one location to another on Earth. Ultimately, the purposes for which spaceports are used in one state may have effects on commerce in multiple states.

I argue that Congress has the power pursuant to the Supremacy Clause to enact legislation designed to satisfy obligations created by the Outer Space Treaty. Congress thus has the power to regulate the commercial space industry to the extent that its actions affect those obligations. Additionally, it has power under the Commerce Clause to enact legislation to satisfy the policy goals of the United States that fall within the purview of the Commerce Clause, giving Congress yet another source of power to regulate the commercial space industry. Thus, while the CSLA is consistent with the purposes of both sources of power to legislate, it is not, in its totality, required or justified by either. I argue that, in the absence of one of these power sources, parts of the CSLA would not constitute valid uses of Congressional power. When considering the possibility of preemption, the deciding entity should determine what set of purposes, and thus, what source of authority, are implicated by a state law that purportedly conflicts with the CSLA. Congress may then carry out the preemption analysis based on the authority most suited to the case at hand, ensuring that the potential preemption is seen in the most appropriate context and that the most appropriate remedy is granted.

---

[209] U.S. CONST. art. 1, § 8, cl. 3.

[210] *See supra* note 182 and accompanying text.

[211] Braniff Airways v. Neb. State Bd. of Equalization & Assessment, 347 U.S. 590, 596-97 (1954).

[212] Outer Space Treaty, *supra* note 1, at art. II.

[213] *See supra* note 182 and accompanying text.

In addition to establishing that a valid treaty exists and gives Congress the power to regulate in order to comply with that treaty, I analyze whether or not, and if so, to what extent, that power gives Congress the ability to preempt state law governing launches and human space flight activity. I therefore proceed to consider each type of preemption, within the context of the larger concerns that unite them all, and the extent to which both the requirements of the OST, as well as national law considered independently, dictate preemption.

A. Preemption Pursuant to the Commerce Clause

i. Express Preemption

The Commercial Space Launch Act, § 70117(c) provides an example of express preemption, the scope of which is explicitly bound by a savings clause. Stating that

> [a] State or political subdivision of a State –
>
> (1) may not adopt or have in effect a law, regulation, standard, or order inconsistent with this chapter; but
>
> (2) may adopt or have in effect a law, regulation, standard, or order consistent with this chapter that is in addition to or more stringent than a requirement of, or regulation prescribed under, this chapter,[214]

this section essentially forbids states from doing what conflict preclusion would most likely prevent them from doing anyway. An express preemption provision creates a presumption against preemption;[215] but exactly what type of state regulation would be inconsistent with this chapter, and whether particular circumstances require preemption under this chapter, is still open to question. In this case, the court must also determine whether state law is more stringent than or in addition to the requirements of the CSLA. The statute is clear about the fact that Congress is instituting minimum standards, but unclear about exactly what those minimum standards are. In order to determine if a law, regulation, standard, or order would be inconsistent with the chapter, the court can look at not only what the chapter says but also what its purpose was. The court might then preempt if, given those purposes, it was "the clear and manifest purpose of Congress" to do so.[216] Because determination of the purpose of particular legislation is also the means by which the court determines if other forms of preemption are appropriate, performing conflict and field preclusion analyses might help determine the purposes of this chapter and thereby clear up the ambiguity concerning what state law is "inconsistent with this chapter."[217]

---

[214] Commercial Space Launch Act, 49 U.S.C. § 70117(c) (2006).

ii. Conflict Preemption

Given current state law, no basis for the first type of conflict preclusion, physical impossibility, appears to exist.[218] The second type of conflict preclusion, obstacle preclusion, is appropriate if state law presents "an obstacle to the accomplishment and execution of the full purposes and objectives of Congress."[219] First, the Court must determine the federal objective for enacting the CSLA. As seen above, the purpose of the CSLA of 1984 was to "complement the launching and associated services . . . available from the United States government," "promote economic growth and entrepreneurial activity," and "enable the United States to retain its competitive position internationally."[220] It was to do so through "stable, minimal, and appropriate regulatory guidelines,"[221] regulating only to the extent necessary to comply with international obligations, to protect the "public health and safety, safety of property, and national security interests and foreign policy interests of the United States," and to increase the efficiency of the licensing process.[222] The CSLA specifically reserved the right of the states to regulate space launch activities within their jurisdictions.[223] The 1988 Amendments instituted the new risk-sharing regime in an attempt to make the United States more competitive internationally; to encourage commercial sector involvement in space launch activities; to encourage the transfer of technology to the commercial sector; and to comply with its international obligations under the Liability Convention.[224] The 1991 NASA Authorization Act expressed an intention to invest in and encourage development of a space transportation infrastructure, as well as state and private investment in that infrastructure.[225] Finally, the 2004 Amendments were intended "to promote the development of the emerging commercial human space flight industry,"[226] balancing safety interests of human space flight participants and crew with the desire to encourage innovation by imposing minimal regulation.[227] Additionally, streamlining the licensing process in order to minimize barriers to commercialization was a priority.[228]

---

[215] Christopher H. Schroeder, *Supreme Court Preemption Doctrine*, *in* PREEMPTION CHOICE: THE THEORY, LAW, AND REALITY OF FEDERALISM'S CORE QUESTION 119, 122 (William W. Buzbee ed., 2009).

[216] Santa Fe Elevator Corp. v. Ill. Commerce Comm'n, 331 U.S. 218, 230 (1947).

[217] *See supra* note 214 and accompanying text.

[218] Physical impossibility preemption occurs when it is impossible to simultaneously comply with both state and federal law. This generally means that state and federal law require mutually exclusive actions. CHRISTOPHER N. MAY & ALLAN IDES, CONSTITUTIONAL LAW: NATIONAL POWER AND FEDERALISM 269 (4th ed. 2007).

[219] Hines v. Davidowitz, 312 U.S. 52, 67 (1941).

Thus, on a national level, the general object of the CSLA was to promote the United States commercial space industry by encouraging innovation and investment; against this general goal, however, was always balanced the desire to protect public health and safety, as well as foreign policy interests.

---

[220] Commercial Space Launch Act of 1984, *supra* note 8, at § 2(4), (5), 3(1).

[221] *Id.* § 2(6).

[222] *Id.* §§ 2(7), 3(2), (3), 5(2), 16, 20.

[223] *Id.* § 21(a).

[224] *See supra* notes 40-46 and accompanying text.

[225] *See supra* notes 58-62 and accompanying text.

[226] Commercial Space Launch Amendments Act of 2004, Pub. L. No. 108-492, 118 Stat. 3974 (codified at 49 U.S.C. §§ 70101-70121 (2006)).

[227] 150 CONG. REC. H10,049 (daily ed. Nov. 19, 2004) (statement of Mr. Boehlert).

[228] H.R. REP. NO. 108-429, pt. 7, at 8. (2004).

For the most part, states' actions do not seem to create an obstacle to the realization of these objectives. In fact, the states appear to have a goal similar to that of the federal government: to promote and benefit from the development of the commercial space industry. By providing special tax incentives, development zones, and other traditional incentives, states exercise local power to advance local interests in a traditional manner. State governments may be more aware of local dynamics and of how to work with local circumstances to further the interests of the commercial launch industry. By allowing the states to legislate in this area, the national government encourages them to invest and innovate, thus sharing the costs as well as the benefits of doing business with the commercial space launch industry. Further, the infrastructure provided by spaceports and the expertise and reliability provided by spaceport authorities minimizes costs for the commercial space industry and maximizes jobs for the local community.[229] Thus, considering that state legislation establishing space ports, space port authorities, corporations, tax exemptions, and other law designed to attract the commercial space industry would not justify even obstacle preemption, it is safe to say that it is "consistent with this chapter" for the purposes of the express preemption clause.

However, the safety aspects of the CSLA's purpose give some cause for concern. The informed consent statutes of New Mexico and Florida seem to fall short of striking the desired "balance between the need to give a new industry a chance to develop brand-new technology and the desire to provide enough regulation to protect the industry's customers."[230] While Virginia explicitly includes the federal requirement that space flight participants be informed of the risks of space flight generally, and of the history of the spacecraft in particular, in its waiver of liability, Florida and New Mexico include only information concerning the scope of the liability waived under state law.[231] Because the only requirement for such a waiver is that that waiver be signed, the statutes of these two states seem to at least implicitly undermine the national requirements, and thus the national safety objective.

---

[229] *See* FED. AVIATION ADMIN., *supra* note 84, at 7-8.

[230] 150 CONG. REC. H10,049 (daily ed. Nov. 19, 2004) (statement of Mr. Boehlert).

Further, without a conversation in which an RLV operator explains "inherent risks" in language that the space flight participant can understand, even the warning that does exist does little to inform a space flight participant about the risks in a way that allows him or her to know the full extent of the danger of the activity and waive claims for damages.[232] As regulations that appear to be less stringent than those required by the CSLA, these warnings seem to violate the express preemption clause's requirement that states enact only more stringent regulation. This state of the law is dangerous not only for space flight participants, but also for the RLV operators that expect to be released from this liability and may be open to suit and liability if preemption occurs.

Additionally, there are some regulations which, if states were to enact them, would almost definitely be preempted by federal law. The first and most apparent of these is a licensing regime. While states could create a licensing regime that supplements but does not directly conflict with the federal regime, to inflict potential licensees with this extra hurdle at all would, I argue, pose a threat to one of the primary objects of the CSLA and its amendments: efficiency. Finally, if states were to adopt more lenient regulation as an alternative to federal regulation, this would not only obviously undermine federal objectives, but also violate the express preemption clause of the CSLA.[233]

---

[231] *See supra* notes 170-73 and accompanying text.

[232] *See* FED. AVIATION ADMIN., DRAFT GUIDELINES FOR COMMERCIAL SUBORBITAL REUSABLE LAUNCH VEHICLE OPERATIONS WITH SPACE FLIGHT PARTICIPANTS 2-4 (2005) (noting how the RLV operator should explain the risks in an understandable way to the space flight participant).

[233] *See* 49 U.S.C. § 70117(c) (2006); *see supra* notes 214 and accompanying text.

B. Preemption Pursuant to the Outer Space Treaty

i. Express Preemption

To the extent that the OST is non-self-executing, Congress is granted the power to legislate in a manner necessary to comply with it provisions. Congress's use of at least some of this power is manifested in the CSLA. The way in which Congress explicitly addresses the relationship between international law and the CSLA indicates that any state law that would be inconsistent with international law, regardless of whether valid under the CSLA savings clause, would likely be invalid as well. In stating that "[t]he Secretary of Transportation shall . . . carry out this chapter consistent with an obligation the United States Government assumes in a treaty, convention, or agreement in force between the Government and the government of a foreign country,"[234] as well as by giving the Secretary discretion to make decisions based on foreign policy concerns, Congress puts the power to regulate for the purposes of the OST squarely in the hands of the executive. Thus, it is plausible that it would be "inconsistent with this chapter" for states to enact law that interferes with obligations of international law as defined by Congress and with the way that Congress, through the Secretary, has chosen to comply with those obligations. However, the precise scope of the minimal standards and consequent obligations to comply with international law imposed by the CSLA is ambiguous. The OST leads to Congressional power to regulate within a limited sphere. However, it is unclear not only what parts of the CSLA are part of that sphere, but also what type of state legislation would be sufficiently inconsistent with those parts to justify preemption. I again look to implied preemption analyses to help determine the purposes of provisions relevant to the OST, and thus the definition of "inconsistent with this chapter" for the purpose of those provisions.

---

[234] 49 U.S.C. § 70117(e) (2006).

ii. Conflict Preemption

The CSLA and 1988 Amendments specifically recognize a desire to comply with international obligations through the regulatory regimes they create and facilitate.[235] The nature of these obligations can be determined by reference to the purposes of the treaties, specifically the OST and Liability Convention, with which they were complying.[236] By taking responsibility for the launching practices of private entities within the United States, Congress appears to be attempting to comply with Article VI of the OST. The federal government requires a launching entity to present information about its launch and payload to the executive for approval, ensuring that it knows what is going into space, who is putting objects there, and how the people and objects that launch and are launched are chosen. This knowledge puts Congress in a better position to fulfill the goals of the OST, including international cooperation and compliance with international law.[237]

One of Congress's primary goals was to comply with the launching requirements, as outlined in the Liability Convention, in order to avoid liability.[238] By including an insurance requirement and adopting the risk-sharing regime, Congress attempted to balance its potential international responsibility for the actions of the commercial sector with its desire to promote the commercial sector by limiting its liability. The specific reference to international law in the CSLA indicates that Congress may have also intended to supply Article VI supervision as to other provisions of the OST and other space treaties in order to facilitate the OST's ideals of cooperation and peaceful use of outer space.[239]

---

[235] *Id.*; *see supra* notes 46-48 and accompanying text.

[236] *See* Outer Space Treaty, *supra* note 1; Liability Convention, *supra* note 46.

[237] *See supra* note 3 and accompanying text.

[238] 134 CONG. REC. 12,113 (1988) (statement of Mr. Roe).

[239] The Secretary of Transportation shall—

> (1) carry out this chapter consistent with an obligation the United States Government assumes in a treaty, convention, or agreement in force between the Government and the government of a foreign country; and

None of the laws enacted thus far by states have infringed upon the international role of the United States as a responsible "launching state."[240] States have done nothing to undermine the United States' liability regime, and it is doubtful whether they could do so in a way that would not lead to a physical impossibility conflict preemption, not to mention substantial public safety issues. Perhaps states are aware of the fact that some kind of liability regime is going to exist, and the federal risk-sharing regime is in the states' interest, supplying an incentive that states themselves would struggle to provide. Similarly, if a state enacted a regulation that allowed launching entities to bypass licensing requirements, such as payload review and determination, then that regulation would be subject to preemption (assuming that part of the purpose of the federal regulations is to supervise domestic space activities).

iii. Field Preemption

There is at least a plausible argument for field preemption pursuant to the power given and obligations imposed by the OST. To determine whether field preemption is appropriate, we must consider whether the legislation is "so pervasive as to make reasonable the inference that Congress left no room . . . to supplement it. . . . Or the [legislation] . . . touch[es] a field in which the federal interest is so dominant that the federal system will be assumed to preclude enforcement of state laws on the same subject."[241] I argue that state law might be preempted pursuant to the OST and its general purposes, as well as pursuant to the OST as manifested in the CSLA. In making this argument, I propose that the CSLA be considered merely a part of the entire field, which might serve as evidence of implied preemption of the field defined by the purposes of the OST.[242]

> (2) consider applicable laws and requirements of a foreign country when carrying out this chapter.

Commercial Space Launch Act, 49 U.S.C. 70117(e) (2004).

[240] *See supra* notes 3, 47 and accompanying text.

[241] Rice v. Ill. Commerce Comm'n, 331 U.S. 218, 230 (1947).

[242] *See infra* fig.2.

As the Court did in *Crosby*, I consider the shared expectations of the contracting parties, as indicated by the negotiating and drafting histories and post-ratification understandings of the parties to determine the scope of the potential field to be preempted and the extent to which a particular situation falling within that scope should be considered for preemption. Assuming that responsibility for non-governmental "national activities" includes responsibility for the actions of completely commercial actors, the United States has the responsibility to ensure that states comply not only with the Liability Convention and OST, but also the Rescue and Return Agreement[243] and Registration Convention.[244] When the treaties were drafted States Parties were intended to suffer the consequences of non-compliance because they were intended to have the responsibility of compliance.[245]

As the commercial space sector and its technology evolve and private parties venture into outer space, international requirements concerning how states must behave in outer space may begin to apply to the actions of non-state actors. The United States as a sovereign in the international field has the ability to extend its jurisdiction into outer space through its people and objects. For a state regulation to extend to objects and people in outer space would be akin to a state regulating a ship on the open seas.[246]

While states can regulate in ways that may have international effects, they cannot do so in a manner that interferes substantially with the national ability to do so, as we see in *Crosby*.[247] Agreement about an end does not justify any means to achieve it.[248] The primary concern of the U.S. civil space program has traditionally been foreign policy. If states begin promulgating their own foreign policies, if only by virtue of their actions in outer space, they will not only create inconsistent policy across the nation, undermining national foreign policy but will also undermine the federal government's sovereignty from an international point of view and thus the ability of the United States as a whole to negotiate with other space-faring nations. It is difficult to imagine a case in which a state law would not undermine a federal regulation intended to ensure that citizens comply with international law, particularly considering the fact that every state would likely want to have its own regulations. The federal interest in this field, due at the very least to the United States' interest in complying with the Liability Convention, may be so pervasive as to justify preemption of the field that gives Congress the power to regulate pursuant to the Supremacy Clause and the OST.

---

[243] Agreement on the Rescue of Astronauts, the Return of Astronauts and the Return of Objects Launched into Outer Space, *opened for signature* Apr. 22, 1968, 19 U.S.T. 7570, 672 U.N.T.S. 119.

[244] Convention on Registration of Objects Launched into Outer Space, *opened for signature* Nov. 12, 1974, 28 U.S.T. 695, 1023 U.N.T.S. 15.

[245] von der Dunk, *supra* note 4, at 28.

[246] *Id.* at 32-34.

[247] *See supra* notes 199-202 and accompanying text.

[248] *See* Crosby v. Nat'l Foreign Trade Council, 530 U.S. 363, 379-80 (2000).

VI. CONCLUSION

The way in which the commercial space industry is regulated is indicative of two purposes: facilitation of commercial growth and innovation and compliance with international law. I argue that these two purposes reflect two, largely overlapping, fields for the purposes of federal preemption. Further, these two fields are regulated pursuant to different grants of federal power: the Commerce Clause and the Supremacy Clause. The power given by the OST via the Supremacy Clause covers a smaller area of the CSLA and currently leads to potential preemption of no more of that law than does that given by the Commerce Clause. However, the fact that the potential preemption of state law by the OST is probably smaller than that dictated by the Commerce Clause might be useful to a state trying to argue against preemption.[249] Additionally, the purposes of the OST represent a field and potentially justify implied field preemption, the scope of which is broader than the CSLA.[250] As the commercial space sector expands, the OST may make it easier to preempt traditional state powers. Further, clear delineation of the purposes and sources of specific CSLA provisions will help judges, relevant executive agencies (like the Department of Transportation), and states attempting to comply with federal law to determine why and to what extent a regulation might be preempted, and to tailor that reasoning to a purpose so as to narrow the scope of preemption as much as possible.

---

[249] *See infra* fig.1, illustrating how each field is defined by its respective purposes and the amount of state law preempted relative to the whole included in each field (which is determined by the type of preemption).

[250] *See infra* fig.2.

At this point, exclusive consideration of federal policy under the Commerce Clause would produce the same potential for preemption in regards to the CSLA in the absence of the OST. However, as the commercial space transportation industry develops, and the private sector ventures into space, the more relevant the OST will be, and the level of prospective preemption may increase. As the commercial space community ventures into space, it is important for its key players to consider the potential national and international consequences of their actions. Consideration of preemption concerns, along with other incentives and characteristics offered by each state, may allow investors to predict the consequences of settling their assets within a particular legal regime in a manner that will ultimately make that investment more valuable.

Meredith Blasingame

Figure 1: Preemption Fields

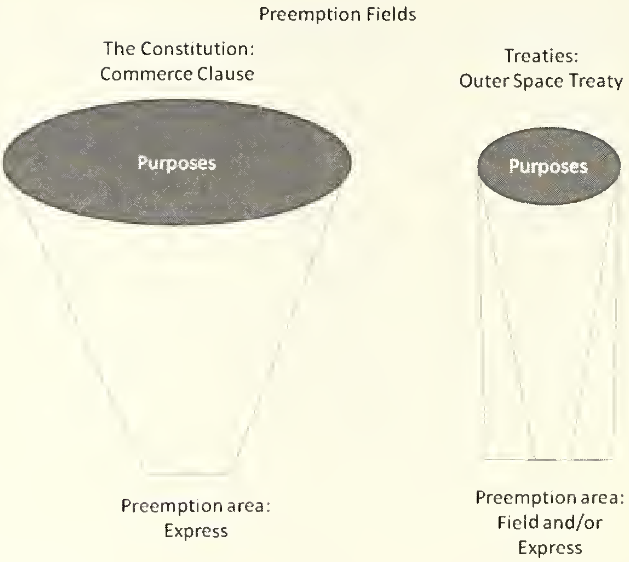

Preemption Fields
The Constitution: Commerce Clause
Treaties: Outer Space Treaty
Purposes
Purposes
Preemption area: Express
Preemption area: Field and/or Express

Figure 1.

Figure 2: Scope and Overlap of Fields

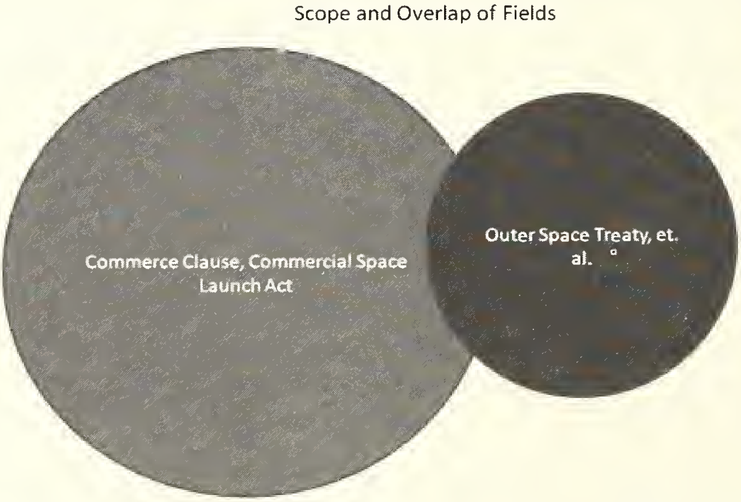

Scope and Overlap of Fields
Commerce Clause, Commercial Space Launch Act
Outer Space Treaty, et. al.

Figure 2.